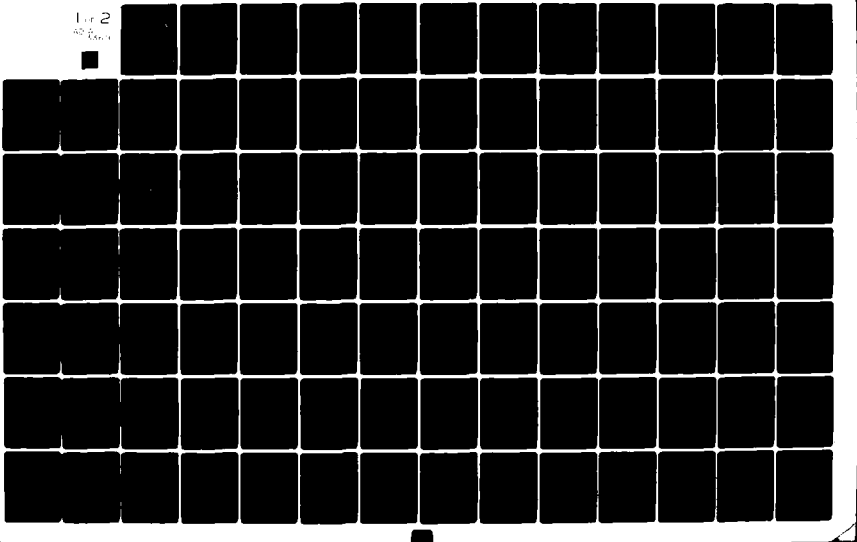
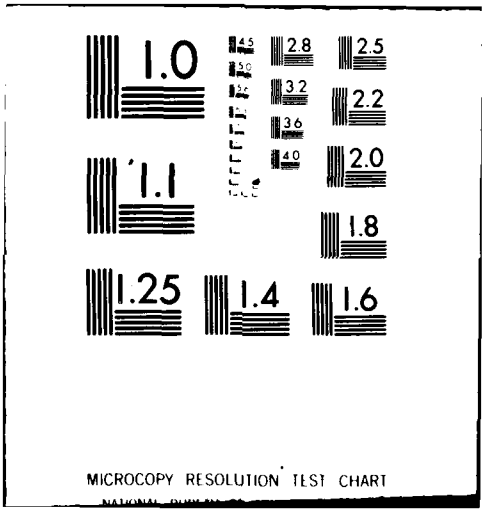
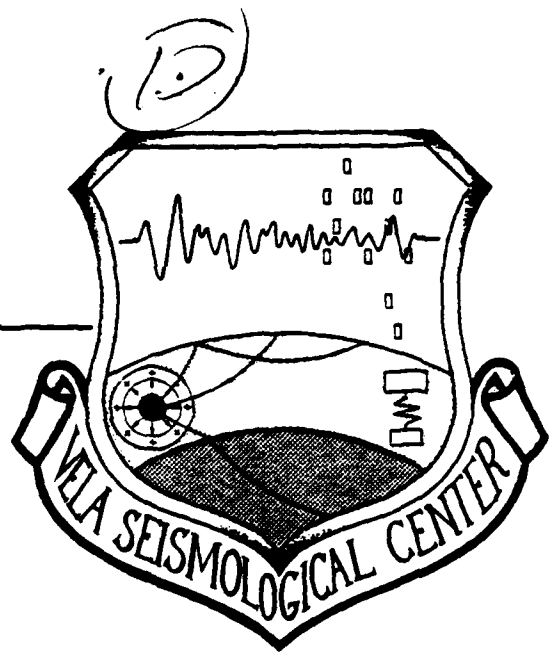
AD-A113 369 SYSTEMS SCIENCE AND SOFTWARE LA JOLLA CA F/G 17/10
AUTOMATIC SEISMIC SIGNAL PROCESSING RESEARCH.(U)
SEP 81 W E FARRELL, J R MURPHY, W L RODI F08606-80-C-0020
UNCLASSIFIED SSS-R-81-5186 VSC-TR-82-17 NL




1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
MICROCOPY RESOLUTION TEST CHART

VSC-TR-82-17

AUTOMATIC SEISMIC SIGNAL PROCESSING RESEARCH



W. E. Farrell

J. R. Murphy

W. L. Rodi

C. B. Archambeau

L. B. Bache

B. Shkoller

SYSTEMS, SCIENCE AND SOFTWARE

P.O. Box 1620

La Jolla, California 92038

FINAL REPORT

September 1981

Monitored by:

VELA Seismological Center
312 Montgomery Street
Alexandria, VA 22314

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER VSC-TR-82-17	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Automatic Seismic Signal Processing Research		5. TYPE OF REPORT & PERIOD COVERED Final Report 6. PERFORMING ORG. REPORT NUMBER SSS-R-81-5186
7. AUTHOR(s) W. E. Farrell, C. B. Archambeau J. R. Murphy, L. B. Bache W. L. Rodi, B. Shkoller		8. CONTRACT OR GRANT NUMBER(s) F08606-80-C-0020
9. PERFORMING ORGANIZATION NAME AND ADDRESS Systems, Science and Software P.O. Box 1620 La Jolla, California 92038		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS AFTAC Project VT/0704 P652497
11. CONTROLLING OFFICE NAME AND ADDRESS		12. REPORT DATE July 1981 13. NUMBER OF PAGES 105
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) VELA Seismological Center 312 Montgomery Street Alexandria, Virginia 22314		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

Approved for Public Release, Distribution Unlimited.

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Spectral discriminants
Narrow band filtering
Multivariate discrimination
Seismic signal processing
Body wave magnitude
Surface wave magnitude
Complexity
Fisher discriminant

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report describes computer programs implemented on the Seismic Data Analysis Center computer system for automatically analyzing seismograms and classifying them as being either explosion-like or earthquake-like. The principal accomplishments of this work have been (1) the development at SDAC of a library of general purpose data analysis programs, (2) the definition of an event-oriented waveform data base (3) the development of a program to measure spectral ratios, body-wave magnitude, surface wave magnitude and

(continued)

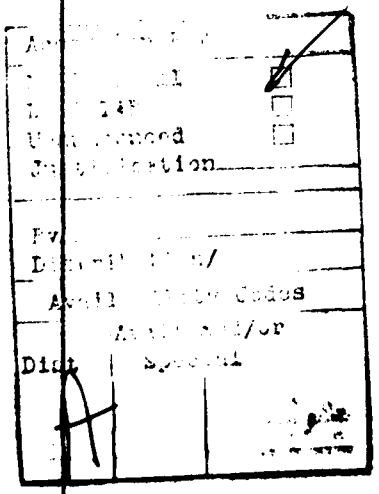
DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Abstract: (continued)

complexity automatically, and (4) the re-analysis of the variable frequency magnitude (VFM) data compiled during the Area of Interest (AI) experiment using linear discriminant analysis. Although successful performance of the discrimination software has been demonstrated in both an interactive and automatic mode, intensive testing and evaluation will require three further steps. The first of these is to incorporate automatic procedures for measuring time-domain m_b and M_s measures. The second is to expand greatly the size of the data base. The third is to process the augmented data set for all discriminants and use these as training data in a more complete linear discriminant analysis.



RE: Classified References, Distribution Unlimited
No change in distribution statement per Mr. Brian W. Barker, AFTAC/VSC.

AFTAC Project Authorization No. VT/0704/B/PMP

Effective Date of Contract: May 15, 1980

Contract Expiration Date: March 31, 1981

Amount of Contract: $96,054

Contract No. F08606-80-C-0020

Principal Investigator and Phone No.

Dr. William E. Farrell, (714) 453-0060, Ext. 257

Project Scientist and Phone No.

Mr. Brian W. Barker, (202) 325-7581

This research was supported by the Air Force Technical Applications Center and was monitored by AFTAC/VSC, Patrick Air Force Base, Florida, 32925, under Contract No. F08606-80-C-0020.

W/O 11145

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS (continued)

LIST OF ILLUSTRATIONS (continued)

LIST OF TABLES

1. INTRODUCTION AND SUMMARY

Discriminating between the seismograms obtained from natural earthquakes and seismograms obtained from underground nuclear explosions is a key procedure in monitoring present and future compliance with treaties limiting the testing of underground nuclear explosions. There is an enormous literature on this problem. A complete discussion of various approaches to seismic discrimination, up to approximately the middle 1970's is contained in the book by Dahlman and Israelson (1977). This book discusses most of the computational procedures which have been proposed whereby seismograms can be anaylyzed in order to infer whether they are earthquake like or explosion like. There has been much progress since this book was written, principally in the testing and comparison of various discriminants. Most studies, however, have been of a piecemeal nature because it is only recently that large digital data sets have been collected together, allowing rigorous comparison of the efficacy of the various discriminants.

Perhaps the most complete examination of methods of seismic discrimination was of the Area of Interest (AI) experiment sponsored by the VELA Seismological Center (VSC) and completed approximately a year and a half ago. For this experiment, seismograms recorded at approximately 30 stations around the world for about 120 events were collected together and distributed to three participants to apply seismic discrimination processing. The results of this test have been extensively reported (Rivers, et al., 1979a, 1979b, 1979c; Savino, et al., 1979, 1980a, 1980b; and Sax, et al., 1979a, 1979b). A review of the findings of this experiment has been provided by Rivers, et al. (1981). Rivers has presented numerous conclusions and recommendations based upon his review of the experiment. They basically fall into two categories. The first are problems associated with the fabrication of the data base (for example, the difference in magnitudes of the earthquakes versus the explosions), and second, discrepencies in computational methodology between the various participants (for example, differences in phases identified for various measurements).

This report describes the results of a study focused on the computational methodology for seismic discrimination. In particular, it describes the development of a set of computer programs for performing automatic seismic discrimination. This study has resulted in the design and the implementation of a bare bones automatic discrimination computer program which makes automatic measurements on seismograms and then, based on prior analysis of training data, classifies the seismograms and assigns probabilities to that classification. The design, operation and performance of this automatic system are discussed in Chapter 2. Section 2.1 of Chapter 2 focuses upon questions of the data base, the measurement of seismogram features (discriminants), and, finally, the linear discriminant analysis of these features to perform classification. The rest of the chapter talks about more advanced seismic phase characterization methods and discusses some of the qualitative aspects of regional and teleseismic discrimination. We note here the problems of formal incorporation of network measurements such as location and depth, which it is difficult to quantify within the available statistical framework.

Chapter 3 discusses the statistical framework which is based upon the Fisher Linear Discriminant. Theory shows that this is the best discriminant for some problems (Gaussian errors and equal covariance matrices). We have adopted it here, not so much on theoretical grounds, but from the principal of parsimony. That is, it is the simplest model which performs the classification. We also have adopted this model for another reason, which is based on our belief that it is inevitable in the study of this problem that one will wish to partition the data (for example, within rather small magnitude ranges or for specific source locations). Partitioning inevitably will entail very small sample sizes; hence, inferences made upon them with complicated statistics will not, in general, be very robust. Although most of the discussion in Chapter 3 has been gleaned from standard texts in multivariate data analysis (Young and Calvert, 1974; Gnanadesikan, 1977), we have added two new ideas in applying these methods to the discrimination between explosion and

earthquake seismograms. The new statistical idea presented here is the use of damping to suppress small eigenvalues after singular value decomposition of the sample covariance matrices. This is closely related to the technique of ridge regression. The other new (to most seismologists) trick which has been added, is the application of jackknifing (leave-one-out) for estimating misclassification probabilities. Misclassification probabilities inevitably reflect the composition of the training data. We find that misclassification probabilities based upon traditional statistics (t-test for example) generally are more lax than the probabilities inferred from jackknifing. We have not, in this study, applied the z-transformation or any other data dependent transformation in an attempt to normalize the measurements for we believe that the jackknife procedure provides more realistic estimates in the real world.

A nice feature of the linear discriminant analysis is that it reduces the multidimensional data space for each event recorded at each station to a single scalar measure, in many cases making it easy to spot outliers or anomalous seismograms. We also have not formally discussed the analysis of variance or the importance functions of the various discriminants. Chapter 3 concludes with remarks on the effects of measurement error, the problem of missing measurements, and some speculations on more robust techniques for estimating means and covariances of the training data.

The last chapter, Chapter 4, discusses a preliminary jackknife study of the variable frequency magnitude (VFM) results obtained by Savino and his coworkers in the Area of Interest experiment. We compare the jackknife results at three specific stations against the bivariate methods used previously to perform discrimination. This section addresses some of the issues raised by Rivers in his discussion of data smoothing problems associated with the Area of Interest experiment.

The conclusions and recommendations which have come out of this study are as follows:

1. The speed at which feature measurement and the jackknife calculation operate is so fast that the application of these methods is limited only by the available data. The hindrance to further progress in automatic discrimination is to be found in the areas of data preparation and the fabrication of homogeneous data sets.

2. The structure of the programs is flexible enough that other discriminants can easily be hooked in to the computational procedures. The results of these new features are easily added to the discrimination data base.

3. The improved statistical methods have only been applied to the VFM discriminant, but similar calculations are now underway on the Geotech data base.

4. We have recognized problems associated with single station versus network data, and problems associated with the incorporation of measurement errors into the analysis. These, however, have not been resolved.

5. A number of known seismic discriminants were not tested in the AI experiment, some of which are traditional time-domain methods and some of which are more advanced waveform modeling methods. It is urged that automatic algorithms be implemented for these techniques and incorporated in the code. This would not only relate the contemporary methods more closely to the older methods, but accelerate the testing of possible future discriminants. Among the methods that fall into these classes are time-domain waveform measurements, depth phases and ARMA models (see Farrell, 1981, and this report, Sections 2.2 and 2.3).

There are three principal research applications for this Automatic Discrimination code. The first is to facilitate the testing of known discriminants on very large (10^4 seismogram) data

sets. It is reasonable to expect that a complete analysis of a data set this size could be done in a man-month. The key to reaching this objective lies entirely in the area of data preparation and data base management. We recommend, for example, that a routine be established now for the regular acquisition of all SRO recordings for every one of the forty or so underground explosions set off each year. The second research application is the use of the code to test new discrimination algorithms. The key to reacning this objective is the writing of an automatic code and its incorporation in the existing package. The third application, and perhaps the most exciting, is to use the code for fundamental studies in regional and teleseismic wave propagation, in particular, path dependent dispersion and attenuation for both the body waves and surface waves. The objective here would be the deterministic modeling of the feature vector using source and propagation physics. The less we rely on statistics, and the more we can apply determinism, the greater our confidence that we can identify sources located in regions for which the historical records are sparse or absent. In our view, this is the real challenge in seismic discrimination.

2. AUTOMATIC SEISMIC DISCRIMINATION

This chapter discusses the structure and operation of an automatic seismic discrimination program package now operational on the PDP 11/70 computer system at the Seismic Data Analysis Center running under the UNIX operating system. Since this project has focused upon the design and the implementation of a discrimination package, very little data has been processed, and no important new results in the area of discrimination per se are presented here. The prime objective has been to fabricate an architecture which will allow the incorporation of a much more complete set of discrimination measurement procedures for a planned extension of the Area of Interest experiment soon to occur. We have currently implemented only the variable frequency magnitude, the complexity, and the surface and body wave magnitude discriminants. Other signal measurement algorithms will be applied as this work continues. These may take the form of alternate methods for calculating traditional discriminants (for example, the various ways of computing the surface wave magnitude), and they may also incorporate the results of current research in advanced methods of seismic discrimination, particularly those which may be applicable to recordings obtained at regional distances. Some preliminary results in this latter area are discussed in section 2.2. Finally, in the concluding section of this chapter, we discuss problems associated with incorporating discriminant measurements which it is difficult to quantify and hence, cannot be incorporated in the current statistical framework.

The procedures discribed here operate by accepting one or more seismograms from a digital data base, analyzing them, and then producing, on a station-by-station basis, an assessment of whether the individual seismograms are more nearly explosion like or earthquake like. This procedure operates with almost no analyst intervention. Essentially, the result of the processing is to reduce each seismogram to a single number, its scalar discriminant -- this number being positive if the seismogram is earthquake like

and negative if it is explosion like. This is clearly a perilous operation, but the application of this automatic discrimination program to a collection of seismograms will not be used in isolation, for there is parallel work underway at Teledyne Geotech in the area of interactive discrimination whereby more searching questions can be asked of the individual seismograms. The purpose of the automatic system is twofold. One, it is to make measurements of features which are thought to contain information about event type; and secondly, on the basis of training data, to classify these features. One important result which accrues from an automatic system such as this is that it will make the searching of voluminous waveform data bases much simpler than it has been in the past. We hope this capability will allow the easy recognition of outliers in the data, that is, stations or events which deviate markedly from past experience. Thus, the final results of this processing cannot be divorced from the training data upon which the discriminant decision is made at the current stage of development. The training data we work with is the Area of Interest data set mentioned previously.

2.1 SYSTEM DESIGN

The automatic seismic discrimination module as currently implemented consists of two computer programs and four data structures. The connections between the programs and the data structures is blocked out in Figure 1. In this figure, we see at the top the first program which is a feature measurement program. This program accepts seismograms from an event oriented data base, recognizes the arrival time of phases within each of the seismograms, and then automatically makes measurements upon those phases. Correctly choosing the time window over which measurements are made is a critical operation for events with body wave magnitude less than about 5.0. For events larger than this, most reasonable event detectors will find the P-wave onset to within a half-second or so; the the surface waves, over most paths, stand well above the background noise for a broad enough frequency band that the

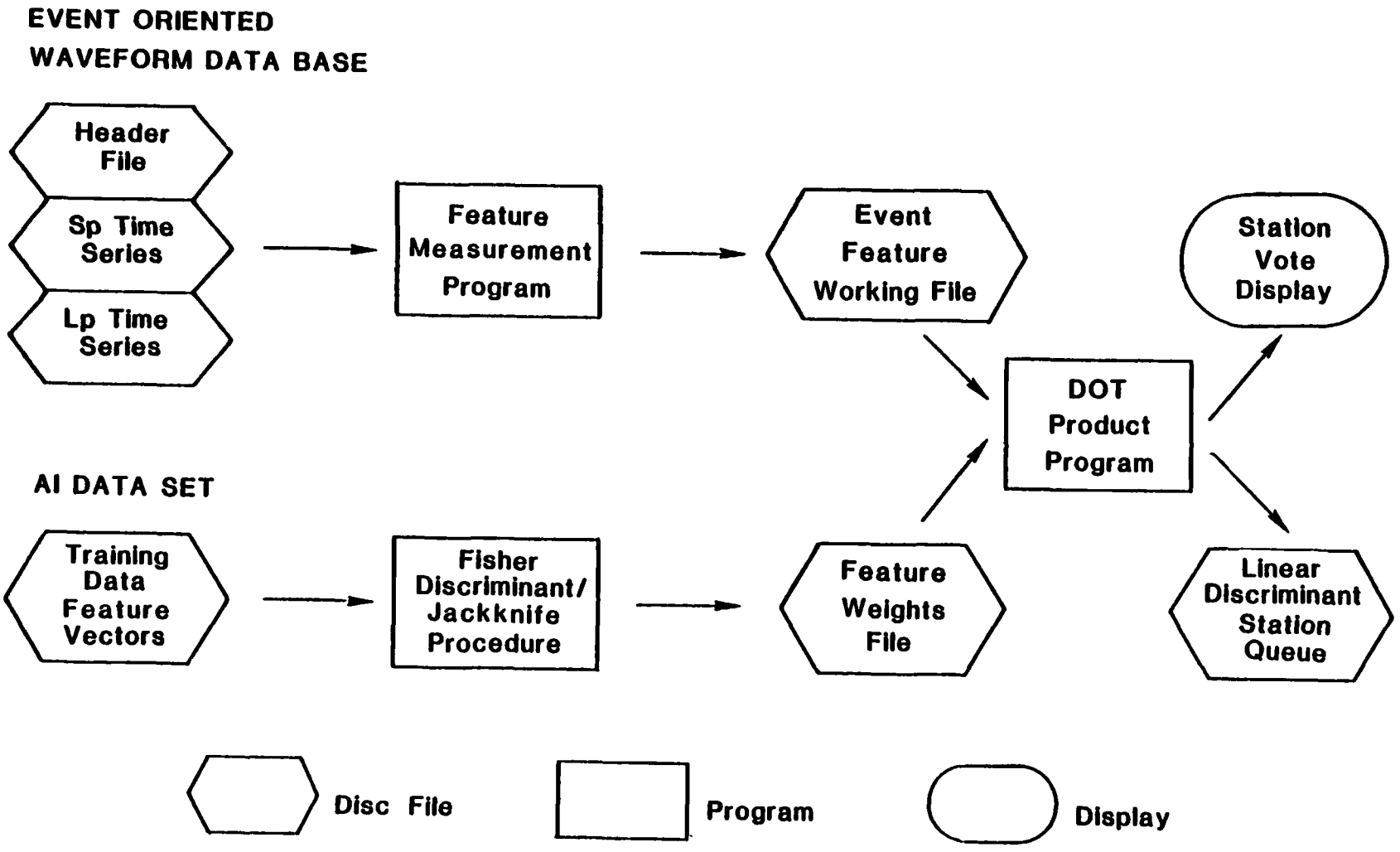


Figure 1. Automatic discrimination is performed by evaluating the dot product between a multi-station event feature file and a multi-station event weights file.

dispersion can be measured. To provide added flexibility in the choice of analysis window in the automatic code, there is a provision to use analyst start time for records which have been visually picked.

The result of the automatic measurements is the creation of an event feature working file which contains a vector for each individual station found in the waveform data base. This feature vector contains, for example, the spectral amplitudes at a set of frequencies; it contains the complexity of a recognized phase; and it will contain other features derived from new algorithms to be added later. Typically, the dimension of the feature vector for each station is of order fifty.

Having located the phase of interest in each seismogram and made the measurements on that phase, the dot product program takes, on a station-by-station basis, a linear combination of these features and evaluates the scalar discriminant for the seismogram. The feature weights file contains a vector for each station. This vector is just the set of weights which have been inferred from a prior analysis of training data.

The procedure for estimating feature weights is shown in the bottom half of Figure 1. It involves creation of training data, feature file, analysis of the file with the Fisher discriminant, and then a jackknife (see Chapter 4) to derive the error probabilities. As more and more seismograms are processed, the event feature working files grow. Eventually this, itself, constitutes a new and expanded data set, and we envisage performing the discriminant and jackknife analysis on this feature file to update the weights to be used to analyze subsequent data which may be processed.

The results of the dot product program are both displayed for each event as it is processed, and also added into a new data base which we call the linear discriminant station queue. Data in this queue summarizes the performance of each individual station over all the events which comprise the total data base. Thus, for every seismogram contained within the data base, we derive a single number which is displayed and queued for further analysis.

The procedures contained within the feature measurement program are outlined in Figure 2. There are four principal elements to this program. The first element is a block which acquires a single seismogram from the data base. The second block processes the seismogram through a comb of Gaussian filters to derive the spectral amplitudes for all pulses occuring within the time window. The phase detector, which is based upon the MARS detector (Farrell, et al., 1980) analyzes the envelope peaks and decides when the phase occurs within the record. The final block in this program performs the feature measurement operation. Not only does it copy the variable frequency magnitude discriminant (Savino, et al., 1980a) to the feature file, it also calculates the signal complexity (Rivers, et al., 1979b, page 33) and measures the spectral magnitudes $\hat{m}_b$ or $\hat{M}_s$ (Bache, et al., 1980), depending on whether the input data was a body wave or a surface wave. It is clear from the structure of this program that additional feature algorithms may easily be incorporated and added to the event feature file.

The other program element in the automatic discrimination package is very elementary (see Figure 3). We call this the dot product program for its principal function is to take a weighted sum of the several feature vectors measured from the various seismograms. The weight vectors are based upon prior analysis of other seismograms recorded at each station. The dot product operation thus reduces each feature vector to a single scalar, the discriminant. In addition to forming the dot product, however, we also provide estimates of the probability that the seismogram at each station arose from an earthquake source or an explosion source. This decision is based upon discriminant means and variances obtained in the course of the analysis of the training data from each station.

2.1.1 Data Base Preparation

The creation of the event organized data base is the most time consuming and, in many respects, the most critical operation in the

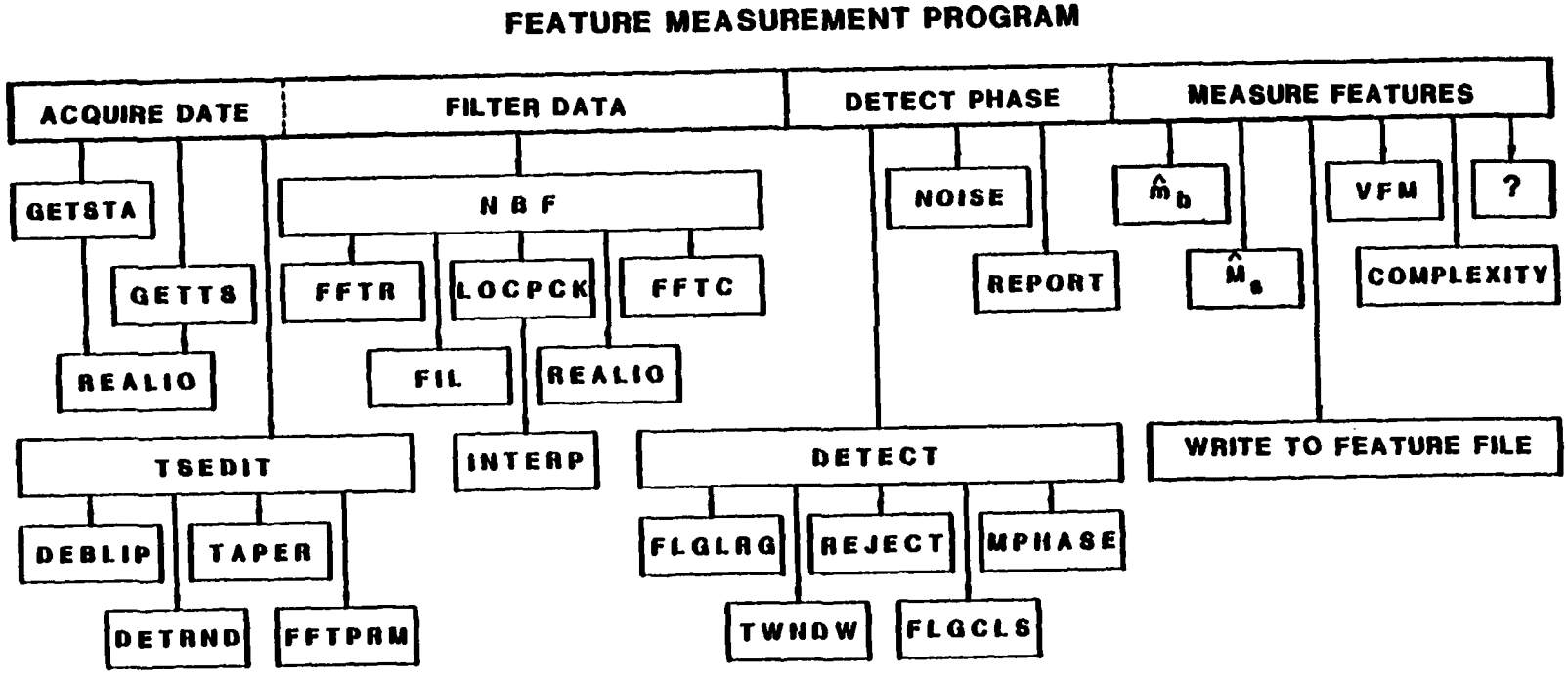


Figure 2. The feature measurement program acquires and edits a time series, detects the phase arrival and measures features of the detected phase.

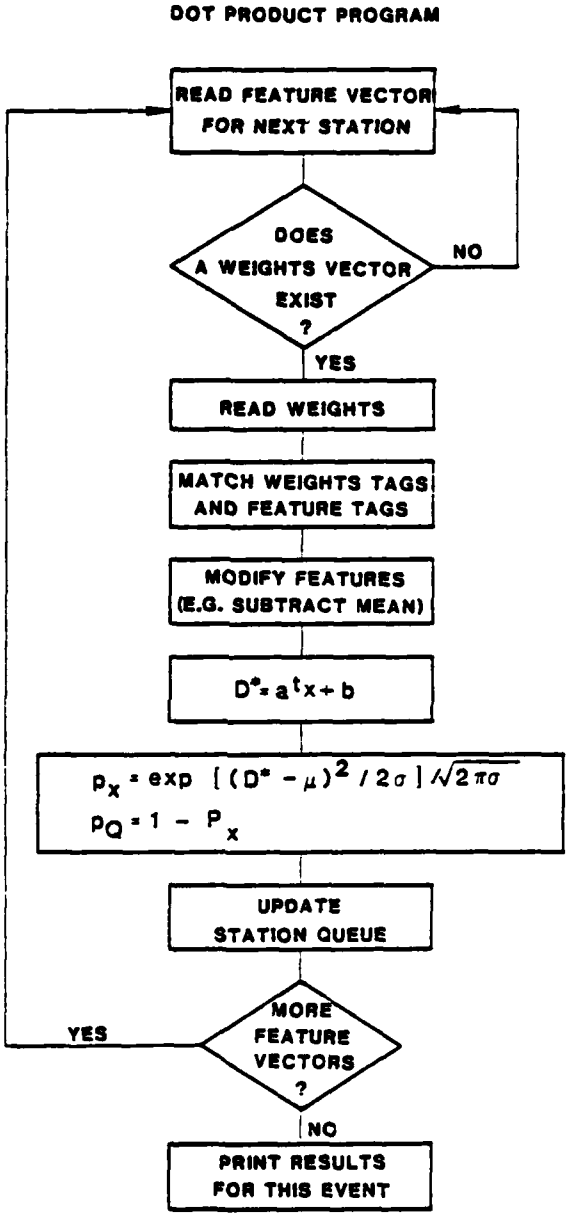


Figure 3. The dot product program evaluates the linear discriminant function $d_j = \underline{a}_k^t \underline{x}_j + b_k$ where $\underline{x}_j$ is a column vector of features from the i^{th} recording of an event, and $\underline{a}_k$ and b_k are the weight vector and constant appropriate for the k^{th} seismic station. The scalars, d_j, comprise the single station estimates of the character of the event.

entire analysis. Data preparation begins with the receipt of subset format data tapes obtained from Data Services at SDAC. The individual seismograms contained in the subset tapes are transcribed into separate time series working files on the PDP 11/70 computer. The subset format header information, giving such quantities as station codes and event origin time, is displayed on the terminal so that the analyst can build the event header file. The data base creation program (see Figure 4) takes the individual time series for one event at the several stations and tailors them into one of two uniform formats, one format applying to the short period time-series and a different format applying to the long period time-series. In the bottom half of Figure 4, we show that there are a variety of interactive display programs which can show the three principal elements in the event waveform file -- either the event header, or the short period time series or the long period time series.

The structure of the data base and its management currently relies heavily upon features contained within the UNIX operating system. This operating system defines directories, subdirectories and files which are linked together in the tree structure outlined in Figure 5. In this figure, if we look at the directory, MVD (for multivariate discrimination), we see this directory is linked to several subdirectories. Reading those links from right to left, a subdirectory is defined for each distinct seismic data type. The one that we have entirely concentrated upon thus far is the SRO data type. This is again linked to subdirectories, each of which pertains to SRO records for a particular event; for example, ev317 as shown in the figure. Finally, grouped in this directory are the five files which actually contain the seismograms for event 317, or the header information for event 317. SPZ denotes a seismogram file which contains all the short period vertical recordings at all the SRO stations for event 317. Likewise, LPZ, LPN and LPE give the long period vertical, north and east seismogram files. Moving to the left in this figure, we see that another directory at this level is the features directory which has branches to the event feature files constructed from the analysis of all the events contained in

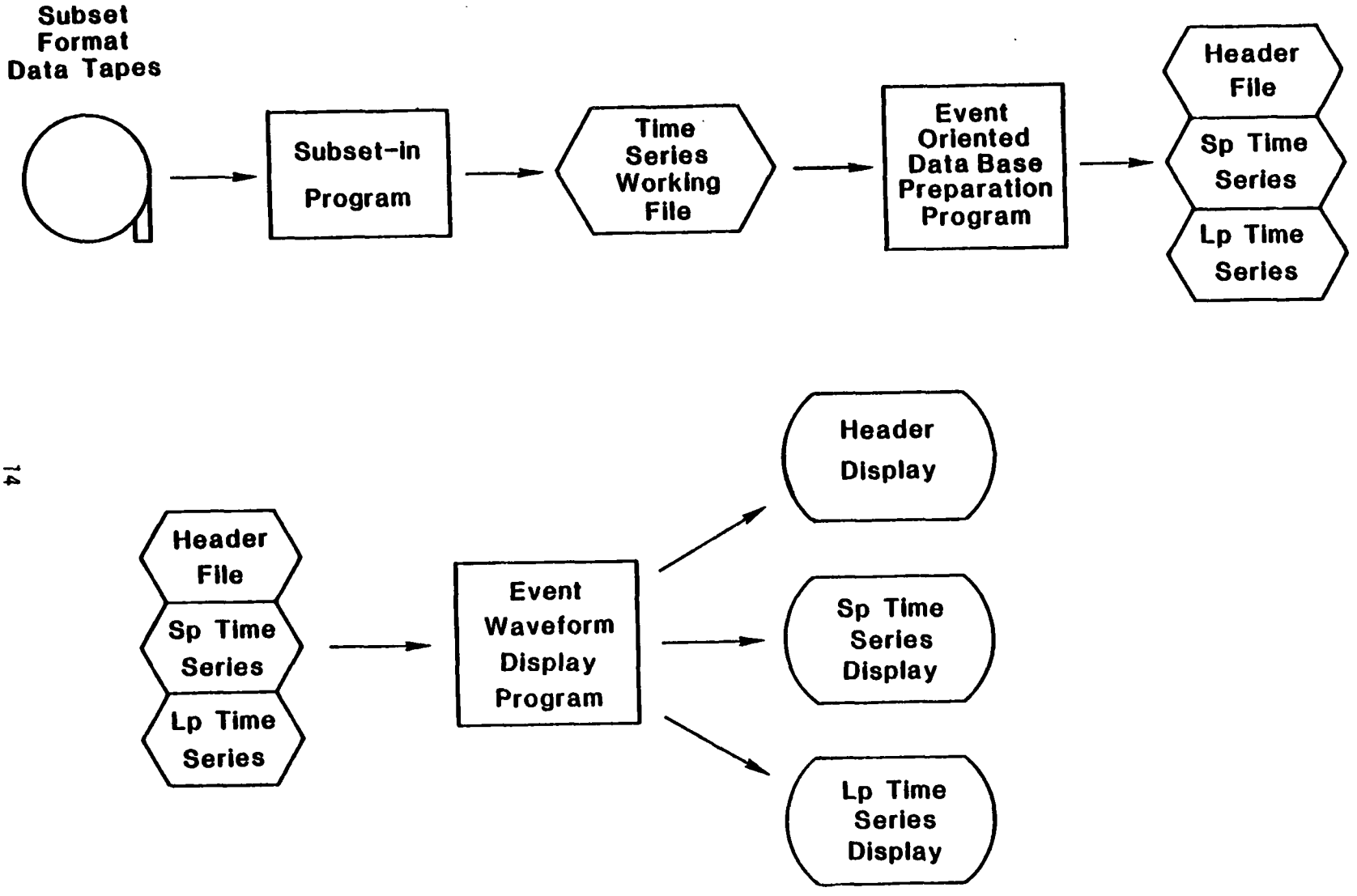


Figure 4. The waveform data base is prepared by creating event oriented data files which are rigidly structured. Graphical displays are an important part of the process.

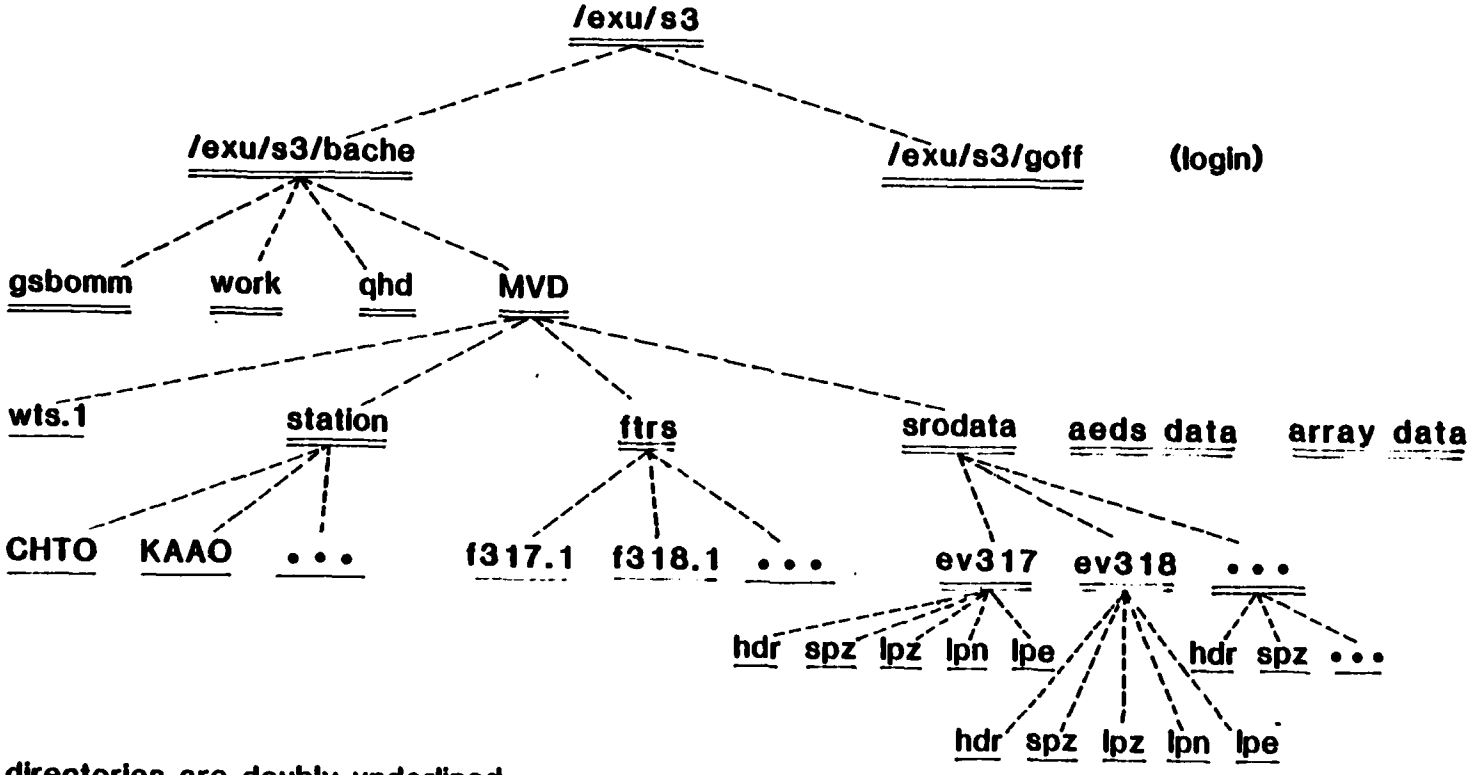


Figure 5. The data files and program files are linked through the directory structure of the Unix operating system. (Essentially identical links are supported by the VAX VMS operating system.)

tne data directories. The station directory shows links to particular station dependent quantities; for example, frequencies and bandwidtns appropriate for the analysis for each station. Finally, tne MVD directory has another branch which points to the weights files used in the dot product analysis.

The makeup of the header file for event 317 within the SRO directory is shown in Figure 6. We see that the file begins with event specific information such as the class, either explosion or earthquake when this is known, the location, the origin time, and otner event related data. Then we see a catalog of all the SRO stations; the code number used at VSC to identify the particular components at the particular stations; the short period and long period sample rates; the start times of either the short period seismogram's window, or the long period seismogram's window; and, finally, the geographic relationship between each station and the event.

Figure 7 shows tne six available short period vertical component seismograms for the SRO stations which recorded event 317. This figure shows a completed plot of all the data contained within the short period vertical file for this event. These short period files are constructed by taking a fixed length of time series, whicn is exactly 50 seconds long. Furthermore, the window for the time-series is tailored such that tne 'expected arrival time' (based on the Herrin tables) for tne P-wave at each station occurs precisely at the fifteenth second within each seismogram. Since the SRO short period recordings are obtained from an event trigger, in many cases the pretrigger data is shorter than the desired 15 seconds, and in all cases the trigger turns off well before the subsequent 35 seconds have elapsed. To fill out tne data, zeros are appended to the beginning and ending of each record to obtain tne predicted body wave arrival at the desired 15 second time. The result of this is to have a rigidly structured file which basically contains a reduced travel-time plot of the seismograms. This shows, for example, that the signal at KAAO has probably been

```
Event: 317   Class: Ex  Location: Shagan River
Lat: 49.987   Lon: 78.949   Depth:    0.0        Mb: 6.0
Date: 28 Oct 79    O-time: 303:03:16:56.9        Ms: 4.18

Sta   Sp# Lp#  Ssr Lsr  S.p. start time  L.p. start time  Dist.    Delta  Azim.

ANMO  000 002   20   1
BOCO  005 007   20   1
CHTO  010 012   20   1  302:03:23:38.0   302:03:23:34.1   3892.8   35.0   337.3
CTAO  015 017   20   1
GUMO  020 022   20   1
KAAO  025 027   20   1  302:02:20:42.3   302:03:12:30.1   1897.3   17.2    22.2
MAIO  030 032   20   1
MAJO  035 037   20   1  302:03:24:53.4   302:03:29:09.5   4900.4   44.1   307.1
NWAO  040 042   20   1
SNZO  045 047   20   1
TATO  050 052   20   1  302:03:24:30.2   302:03:27:23.9   4581.8   41.2   318.5
ZOBO  055 057   20   1
ANTO  060 062   20   1  302:03:23:26.6   302:03:22:45.9   3749.3   33.7    57.0
SHIO  065 067   20   1
KONO  070 072   20   1  302:03:24:14.2   302:03:26:13.7   4372.3   39.3    72.4
GRFO  075 077   20   1

Comments:
%
```

Figure 6. The event header for each sensor type (this is an SRO header) gives important event parameters, sensor waveform windows and event-sensor geometric information. The "comments" field is arbitrarily long and is used to record the processing steps utilized to produce the clean data and important qualitative observations about the individual seismograms.

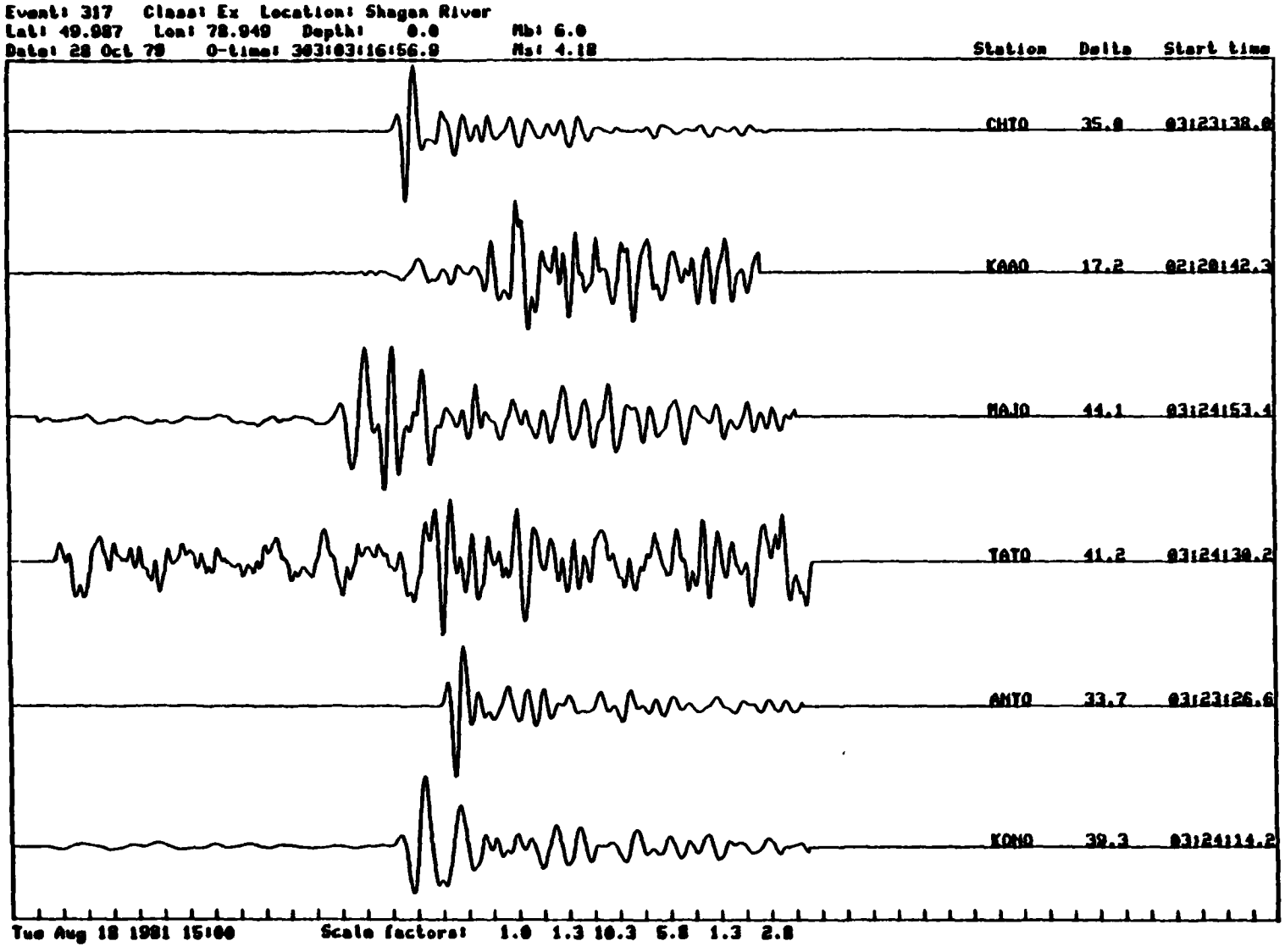


Figure 7. The short period event waveform file for six recordings at SRO stations of a Shagan River explosion shows how the 56 second seismogram window is tailored to start 15 seconds before the predicted P wave arrival time at each site. Tic marks are 1 second apart.

obtained at a regional distance, and indeed the station is only 17 degrees away from the event. Station ANTO has an observed arrival 2 or 3 seconds later than that predicted from the conventional travel-time table. Further, station TATO was noisy during this interval of time, and station MAJO has an arrival 2 seconds or so earlier than that predicted. The scale factors along the bottom show, from top to bottom, the multiplier used to scale each trace to fill the plot window. TATO (5.8) was a particularly weak recording.

It might be thought that the fabrication of such a rigidly structured data base entails unnecessary labor on the part of the analyst before any useful processing can be undertaken. We think, however, that there are many advantages to this procedure. Probably the principal one is that it means that the header file for each event can be much simpler than would otherwise be necessary; it means that the data may be displayed with rather simple graphics programs: it means that the processing can use standard parameters that do not rely upon the erratic start time which it would otherwise be necessary to use; and it means that the scientist examining the seismograms and the results of the processing can maintain in his mind a mental image of what the seismograms look like and where the arrival times occur. All these factors make it easy to control the data quality.

A similar philosophy has guided the construction of the long period seismogram files, an example of which is shown in Figure 8. For the long period seismograms, the available data is tailored so that the window for each station is aligned to place a surface wave traveling at a group velocity of 3 km/sec at the nine hundredth second of the 2000 second record. Again, one can see for station KAAO, it has been necessary to pad the available data with zeros both before and after the seismogram. Station TATO (scale factor 4.5) was again particularly noisy for this event.

The final display (Figure 9) is a geographical plot of Eurasia showing the location of the event and the positions of the seismogram stations which recorded it.

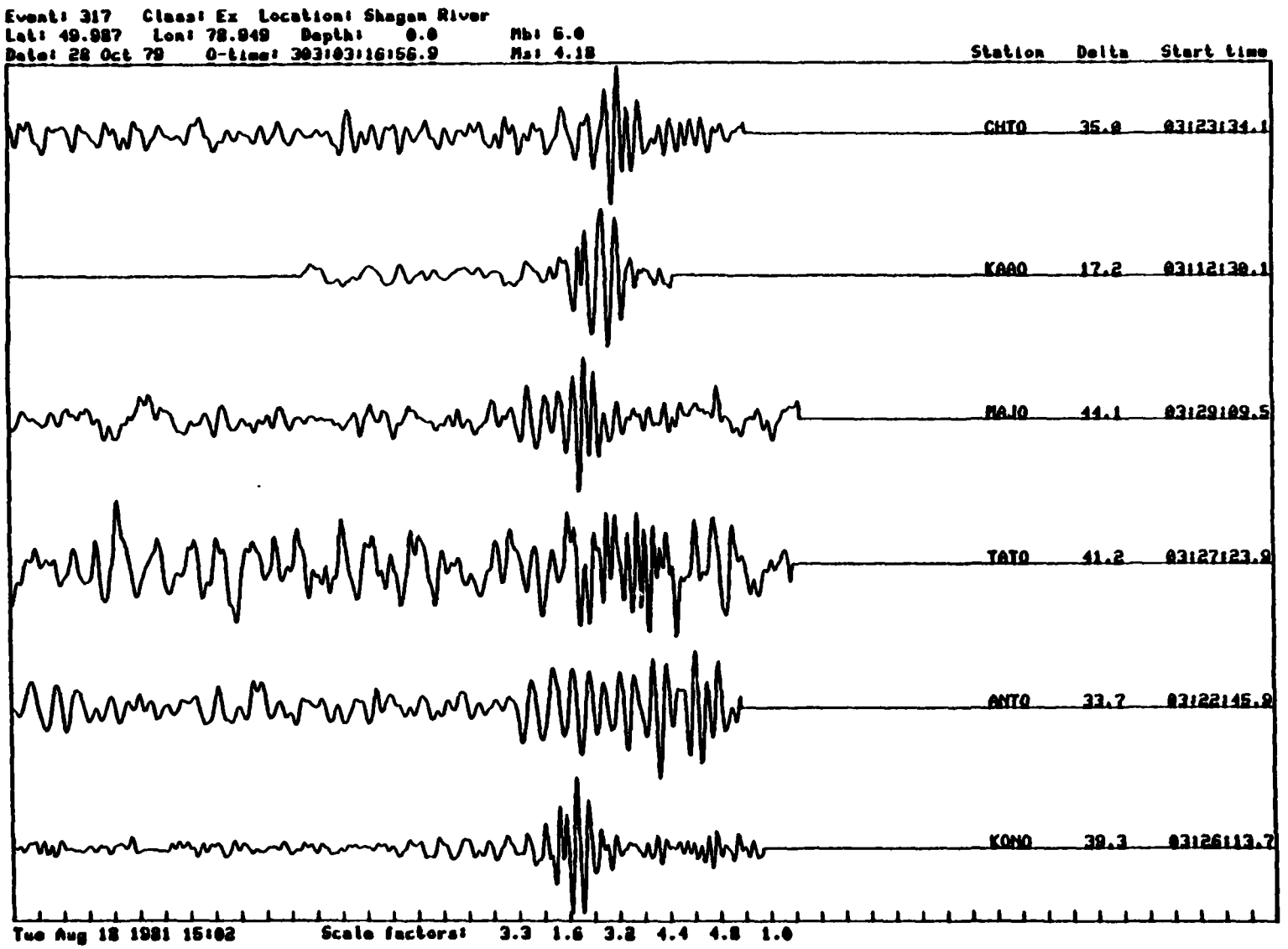


Figure 8. The long period event waveform file for six recordings at SRO stations of a Shagan River explosion shows how the 2000 second seismogram window is tailored to start 900 seconds before the group arrival time of a surface which propagates at a speed of 3.0 km/s. Tic marks are 40 seconds apart.

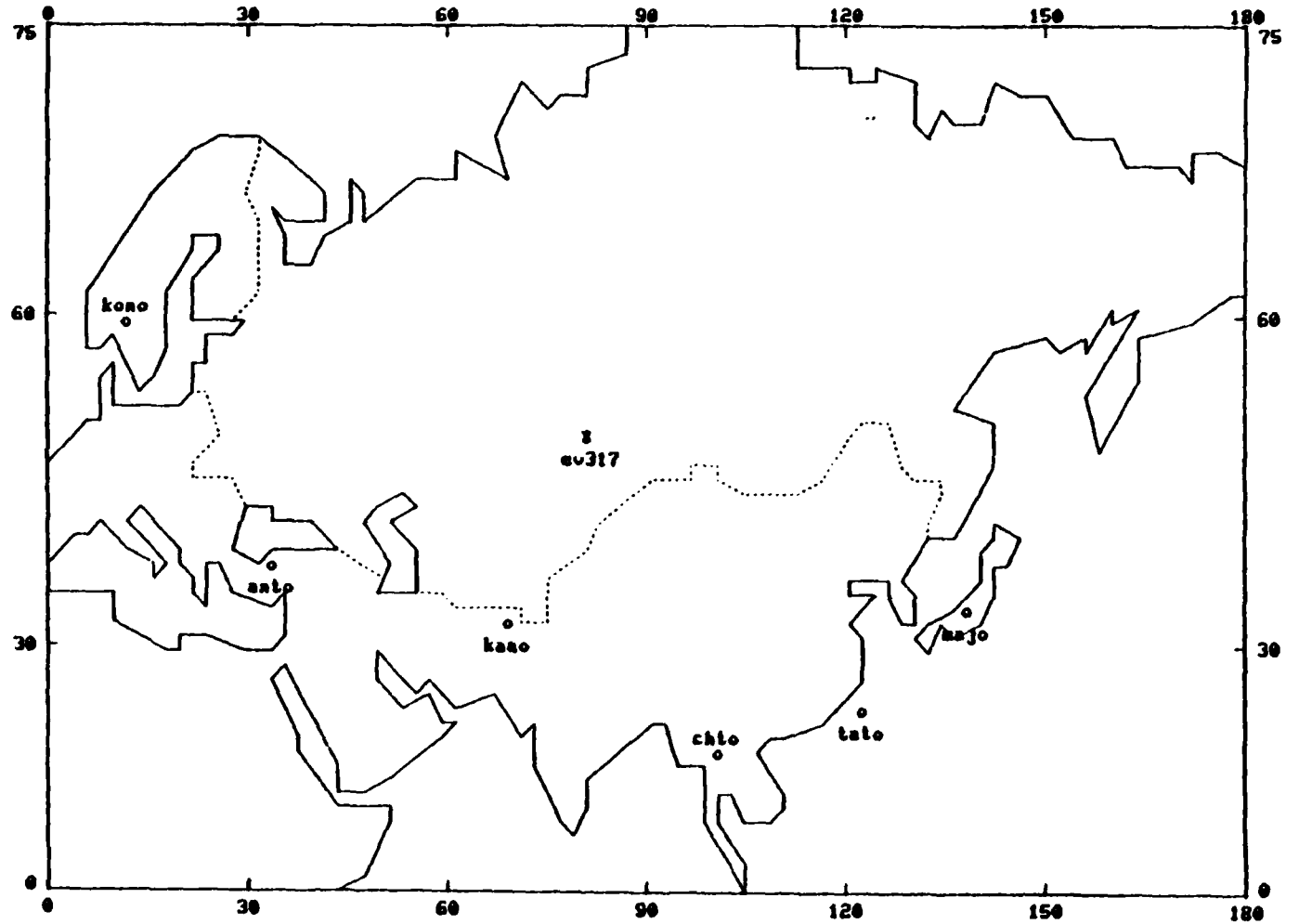


Figure 9. A simplified map of Eurasia is used to display the event location (as contained on the event header file) and the station locations.

2.1.2 Feature Measurement

A great variety of techniques have been proposed for measuring features on seismograms which may yield seismic discriminants. Even among those which are susceptible to automatic analysis, only a few have been implemented in actual algorithms in the current automatic discrimination program. We have elected this approach for two reasons, tne first of which has been the emphasis of this project upon the definition of an automatic discrimination architecture and its demonstration by the actual processing of real data. The second reason is that tne derivation of feature weights must rely upon the existence of a previously processed set of training data and the most accessaDle set of training data available to us has been the variable frequency magnitude measurements reported by Savino, et al. (1980a). It is recognized that many other seismogram features (discriminants) have been proposed and have been studied in more or less detail. We note particularly, the AI lists of discriminants presented by Rivers, et al. (1979a), and Sax, et al. (1979a). Other references for automatic algorithms which we intend to incorporate are given by von Seggern (1977), Chiouris, et al. (1980), and Bache, et al. (1981).

The structure of the feature measurement program, one of the two key elements in the automatic discrimination procedure, was previously shown in Figure 2. Noted in tne right hand portion of that figure are the features (discriminants) with wnich we have obtained actual experience in this project. The discriminants shown tnere as existing in subroutines are tne variable frequency magnitude discriminant, the time-domain complexity discriminant, and the spectral metnods for estimating the body wave magnitude and the surface wave magnitude. Because tne methods of discrimination which we use presuppose the existence of a large set of training data from which discriminant weights can be obtained, the results described in Chapter 4 of this report pertain most particularly to the variable frequency magnitude discriminant, and that is the only discriminant which we are able to process at the moment through the entire automatic discrimination routine.

While developing the feature measurement algorithms, extensive use has been made of visual displays of these measurements in order to check the calculations. A selection of these displays is shown next to demonstrate in more detail the methods of feature measurement and some of the parameters used in the several algorithms. When processing a large body of data, it is expected that these graphical displays of the various features will not be invoked, and, indeed, the display of the features more properly falls in the area of interactive discrimination rather than automatic discrimination which focuses on the end product; that is, the classification of the various seismograms.

Figure 10 shows, at the top, the seismogram for event 317 (Shagan River explosion) recorded on the short-period component of the SRO station at Kabul. Below the seismogram, the narrow band envelope functions are plotted for ten frequencies spanning the range from 0.25 Hz at the top to 4.5 Hz at the bottom. Kabul is only slightly more than 17 degrees away from the Shagan River; so the largest phase picked for this event does not correspond to the first arrival. From the widths of the envelope peaks shown on the various narrow band traces, it can be seen that the time resolution of the filters used to process this seismogram all have a time resolution on the order of one second. The dotted line up the page shows the time at which the automatic detector identified the biggest phase on the seismogram. For frequencies of 2.0 Hz and 2.5 Hz, it can be seen that the phase arrival corresponds to a dip in the spectrum. The actual feature which is measured for these envelope functions is the amplitude of the peak in the envelope either on, or nearest to, the dotted line defining mean phase arrival.

The comb of filters usually used for processing short-period seismograms, while spanning the same frequency range as that of Savino, et al.(1980a), contains ten rather than forty frequency bands and has a somewhat lower Q. The number of filter center frequencies has been restricted in order to limit the dimensions of the feature vectors.

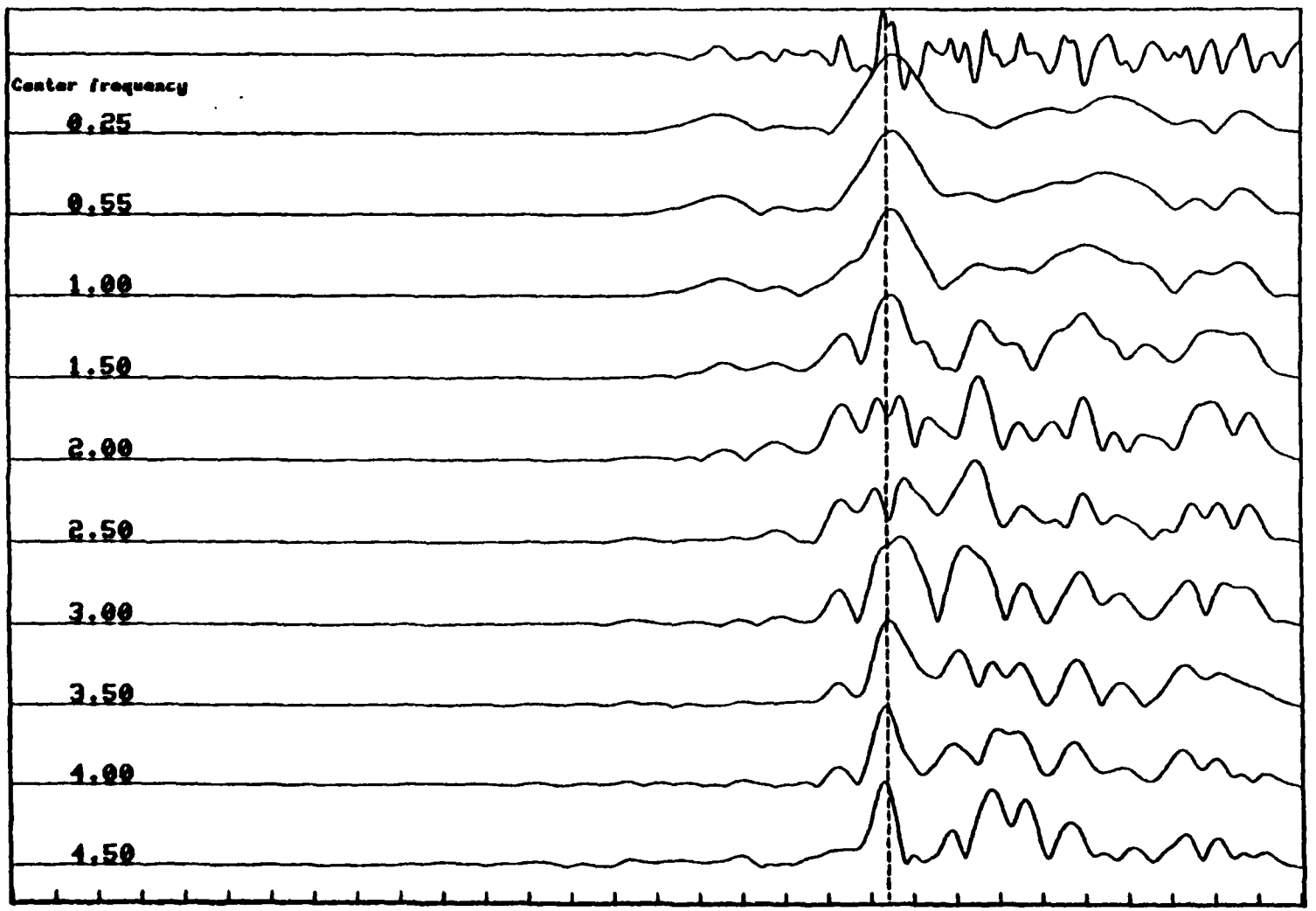


Figure 10. Narrow band envelope functions may be displayed along with the seismogram itself to show the automatic detector phase pick for a short period (SRO) seismogram.

The result from processing the long-period seismogram from event 317 as recorded at Kabul is shown in Figure 11. Although the long-period SRO data contains three channels, vertical and two horizontal, we analyze at the moment simply the vertical channel. Again, the top of this figure shows the raw seismogram. Below that are the ten narrow band envelope functions spanning the frequency range 0.01 Hz to 0.1 Hz, with a dotted line showing the time at which the automatic phase detector identified the maximum signal amplitude. Here, again, the peaks in the envelope function deviate as much as 10 seconds from the mean phase arrival time, and the feature which is measured from these envelope functions is the amplitude of the envelope peak nearest to the mean phase arrival time.

The result of the narrow band filter analysis of the short-period vertical and the long-period vertical seismograms is a set of 20 ground motion amplitudes, ten for each frequency band. These ground motion amplitudes (expressed in nanometers) are corrected for instrument response and then converted to magnitudes. For P-waves, the usual Gutenberg formula

$$m_b(f) = \log_{10}(A(t_p)f)+B(\Delta)$$

is used, where the distance correction is taken from Veith and Clawson, (1972). For surface waves, the formula

$$M_s(f) = \log_{10}(A(t_p)f) + 1.66 \log_{10}\Delta + 3.3$$

is used (Dahlman and Israelson, 1977, page 69). A plot of the ten short-period spectral magnitudes for the Kabul recording of this event is shown in Figure 12. Also shown by the dotted line at the bottom of Figure 12 is a spectrum calculated for a noise window preceding the arrival of the phase. The noise spectrum is defined by the formula

$$m_b^n(f) = \log_{10}(Af) + B(\Delta).$$

where A is the mean envelope amplitude in the noise window.

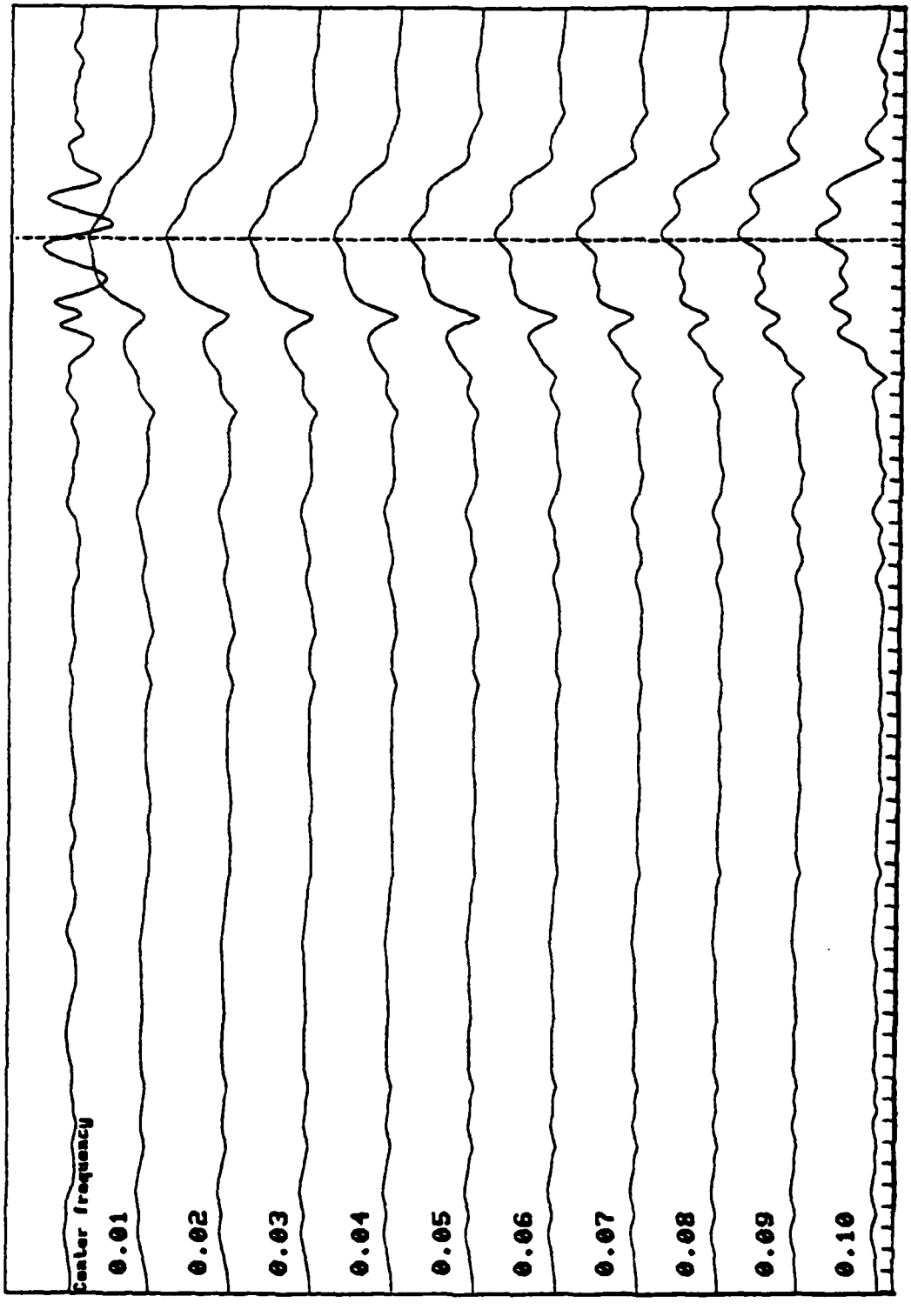


Figure 11. Narrow band envelope functions may be displayed along with the seismogram itself to show the automatic detector phase pick for a long-period (SRO) seismogram.

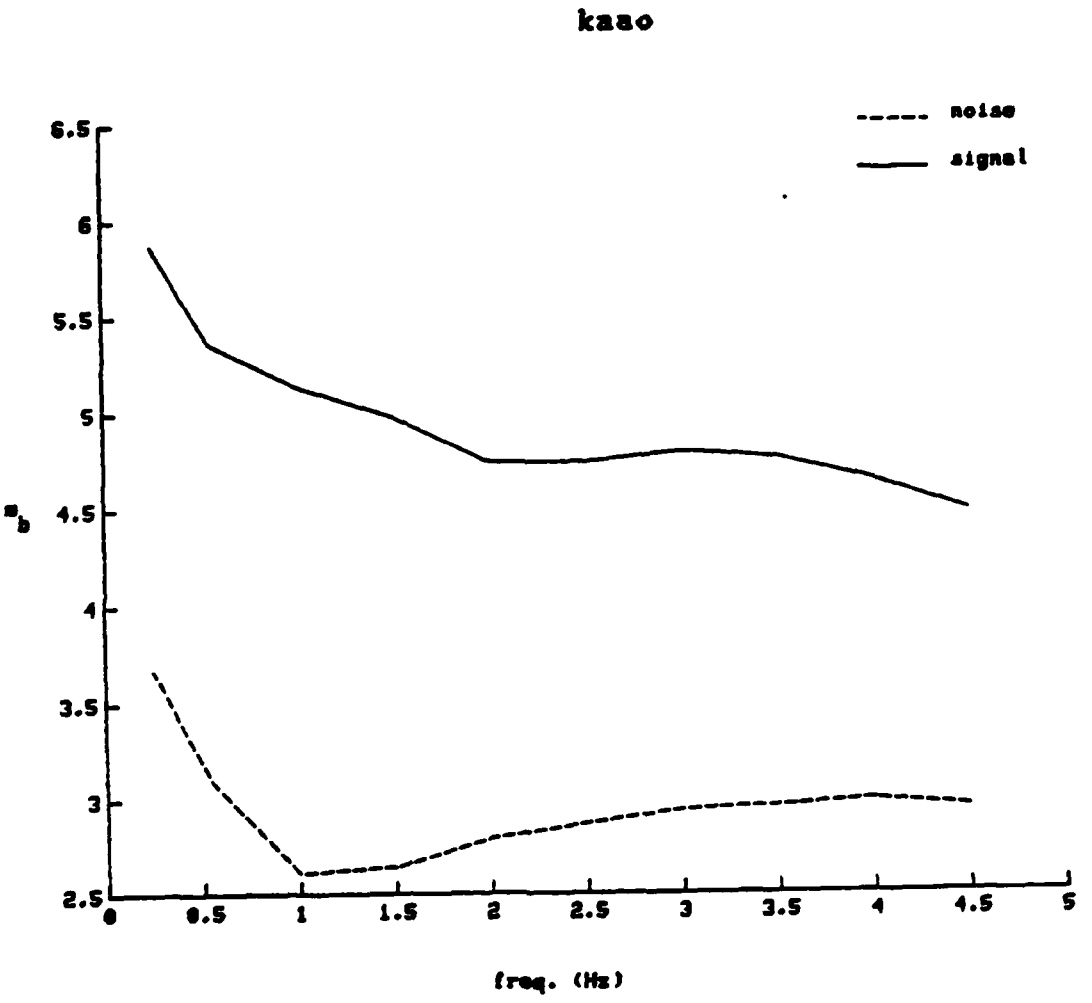


Figure 12. Signal-to-noise ratios may be estimated from plots of narrow band envelope magnitudes and pre-event noise magnitudes. The solid curve in the figure (for the phase identified in Figure 10) is just the VFM part of the event feature vector for station KAAO.

Ten separate long-period magnitudes are measured and written to the station feature vector for each event. It is not possible at present, however, to take the linear combination of these ten magnitudes which best performs discrimination because there does not exist a set of training data from which the discriminant weights can be evaluated. It is possible, however, to make traditional bivariate m_b–M_s plots for each station and each event to compare the short-period and long-period estimates of the spectrum magnitude. An example of this is shown in Figure 13. For making this plot, the simplest possible spectral estimates of surface wave and body wave amplitudes have been used. These are defined to be $\hat{m}_b = m_b$ (1.0 Hz) and $\hat{M}_s = M_s$ (0.05 Hz) with no spectral smoothing or other weighting applied. The slanting line across this figure shows the traditional discrimination relationship $M_s = m_b - 1.0$.

As the quantity of processed data grows, that is, as the feature file becomes larger and larger, we want to compare the new results against all previous ones. Again, this function eventually will fall in the domain of interactive discrimination, but simple bivariate plots of the VFM discriminant are useful for comparing new results against the previous VFM data taken by Savino, et al. (1980a). The example shown in Figure 14 superimposes plots of $\hat{m}_b$ (4.0) and $\hat{m}_b$ (0.55) for event 317 at KAAO on top of the data points calculated by Savino, et al. for Area of Interest seismograms obtained at the same station. It is clear that, upon the basis of these two isolated frequencies alone, event 317 falls within the explosion region of the VFM discrimination plane identified earlier by Savino.

2.1.3 Linear Discriminant Analysis

It was described earlier (Section 2.1 and Figure 3) how automatic discriminantion is effected by taking a linear combination of weignts multiplied by features (the dot product) for each station. The principal end product of this calculation is a display

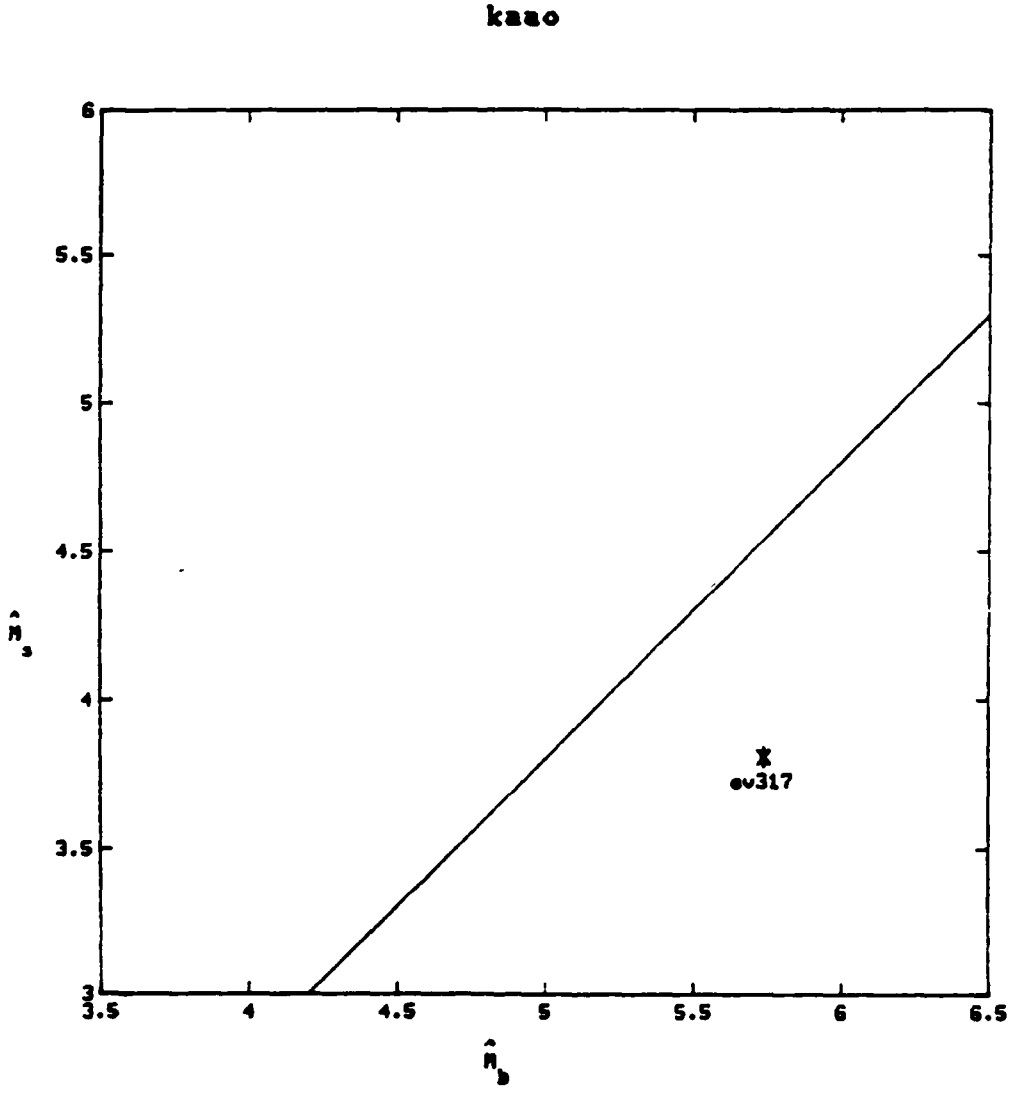


Figure 13. Following feature measurement, bivariate plots of features (in this case $\hat{m}_b$ and $\hat{M}_S$) may be produced.

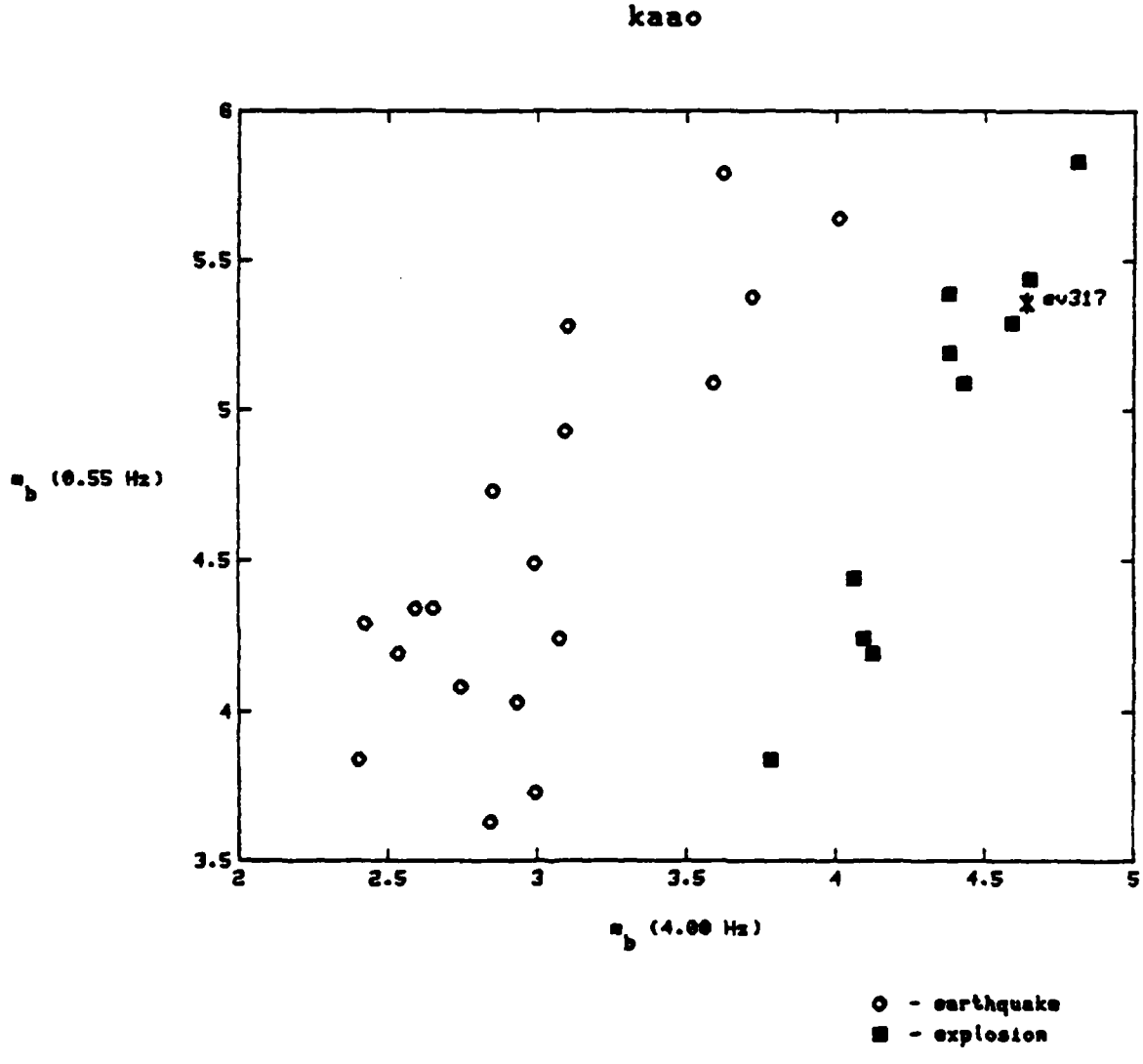


Figure 14. As the event feature file grows, the bivariate plots of features (in this case m_b (high) versus m_b (low)) for many events at a single station may be shown together.

similar to that shown Figure 15. The top part of this figure shows information pertaining to the event. This is simply copied from the event header file and written out to the interactive terminal. The bottom half of the figure shows the results of performing the automatic discrimination for the set of stations listed in the left hand column. For each station, the next two columns give the number of features located in the feature vector, and the number of features for which there are corresponding weights contained in the weights vector. The scalar discriminant d*, evaluated by calculating the dot product of the features for which weights have been found is shown next. (The misclassification probabilities are to appear as two further columns.) At the bottom of the page, the scalar discriminant d* for all available stations is plotted on a horizontal scale ranging from -2 to +2, our convention being that when the discriminant d* is negative, the event is explosion-like, and when it is positive, it is earthquake-like. This plot of d* is similar to those discussed later in Chapter 4 of this report.

Figure 15 shows that when the weights for station CHTO, KAAO, and TATO, as derived from the Fisher discriminant analysis of the Area of Interest VFM data, are applied to a previously unclassified event (Event 317, a Shagan River explosion), it is correctly classified as an explosion at all three stations.

As more data are processed, not only will the feature files for each event and all stations grow, but also the set of discriminant scalars for all stations and all events will accumulate as well (Figure 16). It is planned that these two data sets will be examined in order to identify anomalous events or peculiar stations. These results of the processing are to be used to define a new augmented data set comprising the catalog of feature vectors. They are then to be analyzed afresh in order to derive new feature weights so that the reliability of the discrimination of new events may be improved.

```
Event: 317   Class: Ex  Location: Shagan River
Lat: 49.987   Lon: 78.949   Depth:    0.0        Mb: 6.0
Date: 28 Oct 79    0-time: 302:03:16:56.9        Ms: 4.18

Discriminant file Created:  5 Aug 81        Weights file Created: 22 Jul 81

sta       $raw       $matched     DSTAR        Ex       Eq

CHTO        10           9         -0.38        *
KAAO        10           9         -0.13        *
 No weights for station MAJO
TATO        10           9         -0.35        *
 No weights for station ANTO
 No weights for station KONO

                                    ** *
                  --+---------+---------+---------+---------+--
                  -2.0       -1.0       0.0       1.0       2.0
```

Figure 15. The station vote display (see Figure 1) shows event information (obtained from the header file), creation dates of the feature files and weights files, and the result of classifying the given event.

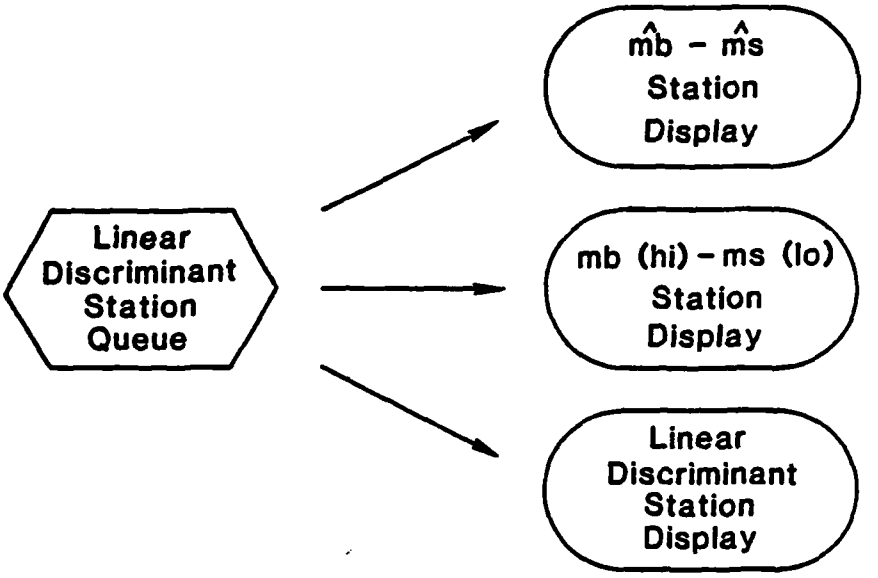


Figure 16. As the suite of processed events grows, the performance of the linear discriminant will be studied on a station-by-station basis by generating station-oriented rather than event-oriented displays.

2.2 TELESEISMIC DISCRIMINATION PROCEDURES

One result that has become apparent from this work is that there is a clear requirement to separate both conceptually, and in the software, the regional discrimination problem from the teleseismic discrimination problem. Several persuasive arguments have led us to this approach. First, of course, much more is known about teleseismic discrimination: algorithms for discriminants can be written down; reasonably good data bases have been collected together; algorithms have been tested in the batch-processing environment during the AI experiment; and, finally, the definition of a "hands-off" automatic code is well underway. Another reason for separating the two discrimination problems is that the short-term objective of this project must be to concentrate on the teleseismic discrimination because of its great impact on the GSS system; yet, not too far in the future, we must be ready with automatic ways of processing single channel or event organized regional signals in case of NSS seismic network deployment. For example, the Regional Event Location System (RELS) with which the automatic discrimination system must be compatible, is focused on regional research, yet initially it will have a data base consisting mostly of teleseisms. Finally, certain practical problems such as association, location and magnitude estimation are performed quite differently for the two classes of signals.

Emphasizing the conceptual differences between regional and teleseismic discrimination and the practical reasons why parts of the software should probably be kept distinct somewhat overstates the polarization that we believe actually exists. A trivial merging of the two problems exists with current operational requirements, for a single event can appear as a regional signal at some stations and as a teleseism at others. This dichotomy was perhaps most clearly apparent in the work of Savino, et al. (1980a) where effective discrimination at Kabul (KAAO) was found to require classifying events by distance, i.e., separating them into four regional or teleseismic categories. It is anticipated that there

will be large areas of overlap between the teleseismic discrimination problem and the regional discrimination problem.

When the data to be processed consists of multiple records from a single event (i.e., associated signals) rather than an isolated waveform, there are two possible ways to proceed. One way presupposes that, as an adjunct of the association process, a valid location has been found. In this case, we can a priori sort the individual traces by epicentral distance and thus classify them as being either regional or teleseismic. Alternatively, we can temporarily ignore the location information, sort automatically, and the ex post facto use a separate associated location/algorithm.

A discussion of the current feature selection algorithms is presented in Section 2.1b. We elaborate here other matters which pertain to Teleseismic Discrimination, discussing specifically: (a) definition of discriminants, (b) feature selection and measurement, and (c) testing and evalution.

(a) Definition of Discriminants

Telesismic discriminants may be separated into two categories, those wnich require prior knowledge of the event location and those which do not need such knowledge, or wnich depend on location knowledge only weakly. Generally speaking, the requirement for location information is equivalent to a requirement for associated signals at three or more widely separated seismic observatories. (Altnough large arrays such as LASA and NORSAR can locate an event with fair accuracy, the smaller arrays in more common use have beams much too broad to furnish more than very approximate location estimates.) In the former category fall the location discriminant itself, depth and network m_b-M_s. In the latter category fall complexity and several frequency domain discriminants, including VFM, or spectral ratio, automatic m_b and M_s and higher moments of frequency.

Location and depth are peculiar discriminants, the use of which in the context of automatic signal processing is not clear at

this time. Although automatic phase identification and timing is necessary for automatic location (and depth estimation), it is certain that location and depth must be treated quite differently from other discriminants. The crux of the matter is that the problem of bias in the training data is paramount for these quantities. Furthermore, depth, when available, is either of overwhelming significance (e.g., deep (100 km) earthquakes can look explosion-like by all the usual measures) or is irrelevant. Thus, knowing that an event is shallow, say less than 35 km, is useless for discrimination. Location is peculiar in a different sense. There are large parts of the area of interest in which neither earthquakes nor explosions have occurred. Suppose now a new event is found to locate in a previously silent region. Is that fact taken alone of any use in discrimination? Conversely, suppose an event is found to occur in, or near, a known test site; is that information alone useful for deciding whether the event is an explosion or earthquake? In both these instances, the location information might be the key that intensive analysis is warranted, but it seems not to help answer the discrimination problem.

It is absurd, of course, to suggest that location is irrelevant for discrimination (although indeed some analyses in the AI discrimination experiment did ignore location). The clear way location enters is through the source and station regionalization of discriminants. An example of source regionalization is the problem of Lake Baikal. An example of station regionalization was the discovery by Savino, et al. (1980a) that, for the VFM discriminant to work most effectively on the AI data set, it was necessary to select a distinct pair of separation frequencies for each station, and for a given station there is some evidence that the separation frequencies depend on epicentral distance.

The mathematical way of expressing this geophysical phenomenon is to say that our data (discriminants from many events at many stations) do not come from a single homogeneous population, or rather two populations, earthquakes and explosions, but instead,

from a multiplicity of populations, each with its own variance-covariance structure. It is a well-known principle in statistics that if you mix good data together with bad data, the bad data always dominates. The practical illustration of the principle in the context of seismic discrimination is the comparison of VFM scatter plots for single station measurements versus network averages. Based upon network average VFM scatter plots, one would be tempted to dismiss the method because of the large overlap in the explosion and earthquake populations. It is only when the VFM discriminant is studied on a station-by-station and source-region by source-region basis that its power emerges. It is quite possible that tnere is a similar hidden structure in the complexity discriminant.

(b) Feature Selection and Measurement

Feature selection (Calvert and Young, 1974, p. 224) is the word used in mathematical statistics to connote the first (and often empirical) step in the hierarchy of operations whereby one distills an enormous quantity of data down to a few bits of information. Oftentimes tne process of feature selection is guided by intuition, or ancillary information. For example, we have a physical reason for supposing that earthquakes might be more complex than explosions, or that explosions generate less surface wave energy than earthquakes of the same bodywave magnitude. These are features which, if it is plausible, ought to be selected for further study. The purpose of feature selection is to reduce the size of the data space so that exotic numerical calculations are possible. One hopes first tnat by combining the features together, one can improve the performance of the discrimination procedure with poorer quality data (i.e., lower magnitude). Furthermore, the calculations required to assign classification probabilities from analysis of numerous training data are computationally unfeasible without feature selection.

We follow the narrow definition of terms common in the statistics texts and distinguish carefully between feature selection and feature extraction. Feature extraction we apply to the next hierarchical procedure where some method, for instance, principal component analysis, is used to effect a further reduction in dimensionality, but based upon rigorous mathematical procedures rather than qualitative or semiquantitative procedures. For example, with reference to the AI experiment, the feature selection (or measurement) part of the analysis was the procedure of choosing and calculating a number (between 20 and 40) of parameters which were thought to contain the useful discrimination information in a 100 to 500 term seismogram. The process of feature extraction then showed that between two and four of the selected features (or linear combinations of them) were sufficient for reliable discrimination. For the VFM method, the selected features were 40 narrow frequency band magnitudes of the P-phase, and the extracted features were the two frequencies (after polynomial smoothing) for which discrimination worked best. Likewise, the other studies used a different selection of features but showed that just three or four linear combinations hold most of the variance in discriminants.

The method which has been implemented for automatic feature selection relies heavily on the QHD processing of individual seismograms. There are three principal steps in the analysis (see Figure 2). The first of these, called Filter, consists of a data edit task (TSEDIT) and a multiple narrow band filter (NBF) task. The second step, Detect, is a phase identification task which picks the arrival time of the event. The performance of the current detector has been described by Farrell, et al. (1980), but the advanced phase detection algorithms described above have not been so exhaustively studied. Finally, the Feature Selection procedure consists of further refinement of the frequency domain discriminants (for example, applying instrument response correction, or converting signal spectral amplitudes into magnitudes), and calculation of time domain discriminants and storage of these features on an event discriminant file.

(c) Testing and Evaluation

Whereas feature selection and automatic waveform processing are at a relatively advanced state of development, the multivariate statistical procedures are at a more rudimentary stage. The mathematical statement of the problem, as we currently see it, is described in Chapter 3. In casting those equations into the geophysical context of discriminating earthquakes from explosions, we have identified three potential problem areas which it is felt should govern the testing and evaluation of the automatic discrimination system. These potential sources of difficulty are: (1) accounting for missing and erroneous data; (2) combining single station discriminants into network (or event) averages; and (3) the size and availability of the training set.

The problem of missing and erroneous data is a very practical one which we do not yet know how to treat mathematically. We are lead to consider erroneous data from the following argument. As the event size decreases, it seems reasonable that the process of feature selection becomes less precise because the signal-to-noise ratio in seismograms themselves degrades. Phases become misidentified, holes appear in the spectrum from noise interference, the complexity measure sees less and less signal, but more and more noise, and surface wave magnitude disappears entirely. What are the implications for discrimination, and how does one quantify this benavior? It is certain that simply associating a standard error (based, for example, on the noise in each record) with each feature is not sufficient because we do not know how to use this information. What we would like to assume, perhaps, is that the covariance in the feature vector consists of two parts, a measurement noise part and a geophysical noise part; and, further, that the measurement noise part can be objectively estimated from a single record, whereas the geophysical part requires a multitude of events and stations.

Another place that missing and erroneous data (it is useful to think of missing data as ordinary data with absurdly large error bars) affects the automatic processing is the case when one station may not report any data from a given event. For example, take a source near the Caspian Sea and assume Kabul is not functioning. Then, it is known from the AI experiment (Savino, et al., 1980a) that Kabul was particularly powerful at discriminating for this source region. However, with the best station now missing, how is it best to treat the data available for the particular event in question? We suspect that the best answer would be to reassess the entire historical data base, calculating a unique discriminant function for the subset of the historical data which best matches the event in question. To throw Kabul out in this hypothetical example would mean a massive reevaluation of the training data which is clearly possible in the off-line (or interactive) environment, but not realistic for the automatic program package. One can easily imagine less disastrous cases of missing data (for example, holes in the spectrum at particular frequencies), but again, quantifying the impact of this phenomenon on discrimination and probability assessment is not yet understood.

A second problem area is the method of joining single station features into a network discriminant in the case where associated signals are being processed. Suppose there are n stations and each of them supplies a single seismogram from which we measure m features. One, first of all, could lump everything together and perform discrimination on a single n x m dimension feature vector. To effect discrimination and assign probabilities, one would have to process the training data similarly, and this would entail numerous costly computations. At the other extreme of dimensionality, one could average each of the m features over all n stations, perhaps as network magnitude is now usually found, following the method of Ringdahl. The clear difficulty here is that the station dependence of discriminants is ignored, and we know from the AI discrimination results that a less powerful test results. The golden mean, we feel, lies in a two step procedure whereby an n + m dimension

problem is solved in stages. First, each station is processed separately, yielding n "votes" as to the event type. The individual station votes then must be combined to get an overall discriminant. We do not yet know how to count the votes in an optimum sense, nor do we know how to attach a probability statement to the final decision.

The third and perhaps most important problem area, and the one that affects testing and evalution most directly, is size and availability of the training data. Practically, it is the small events (say $m_b < 4.5$) which are of most concern to automatic discrimination because of anticipated treaty limitations and the fading out of the m_b-M_s discriminant. This is the event range which is particularly poorly represented in the current AI data set. Although events in the required yield range are rare, there is data which ought to be collected together. Another problem relating to training data is the use of array information. There is no question but that it must be incorporated; yet this cannot yet be achieved because it is not known how to beam small arrays without incoherently attenuating the high frequency part of the signal spectrum. This is a well-recognized problem and is susceptible to solution, but better beaming procedures must be implemented before the array data can be utilized effectively for automatic discrimination. One further point pertaining to the training data is the quantity of historical data which is to be available on-line to the automatic discrimination processor. When the data set becomes large enough, one wants to conduct a variety of discrimination experiments using different partitionings of the data; for example, by magnitude, by source region, by path type, etc. These will be done by making one complete pass with the feature selection code, but many different combinations of feature vectors will be taken for the statistical analysis. Thus, we recommend that further multivariate statistical research be conducted in an off-line mode, and that the automatic processor use the results of that research by simply applying a predefined algorithm to the discriminant vector. One reason for recommending

this procedure is that there will be available an interactive discrimination system as part of RELS, and we feel that this is the more effective way to search the archive and fine-tune the multivariate statistical analysis.

3. AUTOMATIC MULTIVARIATE STATISTICAL DISCRIMINATION

3.1 THEORETICAL CONCEPTS

Our current perception of automatic methods for seismic discriminaion is based on standard statistical approaches similar to those summarized by Tjostheim (1981) and discussed in many statistics textbooks (e.g., Young and Calvert, 1974; Patrick, 1972; Rao, 1973). We highlight here some of the concepts underlying the linear discrimination algorithm discussed in subsequent sections.

The seismic discrimination problem can be posed statistically by treating the various discriminants measured from an event as components of a vector random variable $\underline{x}$, which is called a discriminant vector (also feature vector or pattern vector). The M components of $\underline{x}$ may include any available measurements, including dissimilar quantities (e.g., m_b-M_s, complexity, VFM magnitudes at various frequencies) or a mixture of individual station data and network averages.

Given a measurement of $\underline{x}$ from an unidentified event, the discrimination problem is to infer the event's class, C. C can take the values C_1 (explosion class) or C_2 (earthquake class). The inference of C must ultimately be based on information about the multivariate probability densities of $\underline{x}$ conditioned on the two classes of events: $f(\underline{x}|C_1)$ and $f(\underline{x}|C_2)$. The mean of $f(\underline{x}|C_1)$, for example, describes where in M-dimensional space the discriminant vectors from explosions are expected to fall. Its second and higher moments describe the expected variability (scatter) in the data, such as that caused by inherent differences between events, variations in earth structure, and measurement errors.

When $f(\underline{x}|C_1)$ and $f(\underline{x}|C_2)$ are not known, they must be estimated - whether explicitly or implicitly - from training data sets. These are the discriminant vectors observed from past identified events of each class. The explosion training set will be denoted as the set of vectors $\underline{x}_{i(1)}$, $i = 1, .., N_1$, and the earthquake training set as $\underline{x}_{i(2)}$, $i = 1, .., N_2$.

The class of an event cannot, in general, be determined with 100 percent certainty; so a solution to the discrimination problem must be a statistical statement. A variety of ways of expressing one's uncertainty about C are possible. One is in the form of "posterior probabilities" that the event belongs to C_1 or C_2: $P(C_1|\underline{x})$ and $P(C_2|\underline{x})$, respectively. The adjective "posterior" refers to the fact that the probabilities are determined after $\underline{x}$ has been measured. Posterior probabilities require the assumption of <u>prior</u> probabilities of C_1 and C_2, $P(C_1)$ and $P(C_2)$, which anticipate the relative likelihood of each class before the event has occurred. Normally one would set $P(C_1) = P(C_2) = 1/2$. Bayes' Rule gives the posterior probabilities as

$$P(C_1|\underline{x}) = \frac{f(\underline{x}|C_1)P(C_1)}{f(\underline{x}|C_1)P(C_1) + f(\underline{x}|C_2)P(C_2)} \qquad (1)$$

$$P(C_2|\underline{x}) = 1 - P(C_1|\underline{x}) \quad .$$

Interpreted literally, this type of solution does not classify the event, but simply describes how earthquake-like versus explosion-like the event is.

A second type of solution is an actual classification based on a decision function $D(\underline{x})$. D takes scalar values and assigns a class $\hat{C}$ to an event, according to the rule

$$\begin{aligned} \hat{C} &= C_1 \quad \text{when} \quad D(\underline{x}) < 0 \\ \hat{C} &= C_2 \quad \text{when} \quad D(\underline{x}) > 0 \quad . \end{aligned} \qquad (2)$$

The equation $D(\underline{x}) = 0$ describes an (M-1)-dimensional hypersurface which divides discriminant space into "decision regions" $\mathcal{R}_1$ and $\mathcal{R}_2$ (see Figure 17). Equation (2) thus tests whether $\underline{x}$ falls in $\mathcal{R}_1$ or $\mathcal{R}_2$, and so is equivalent to

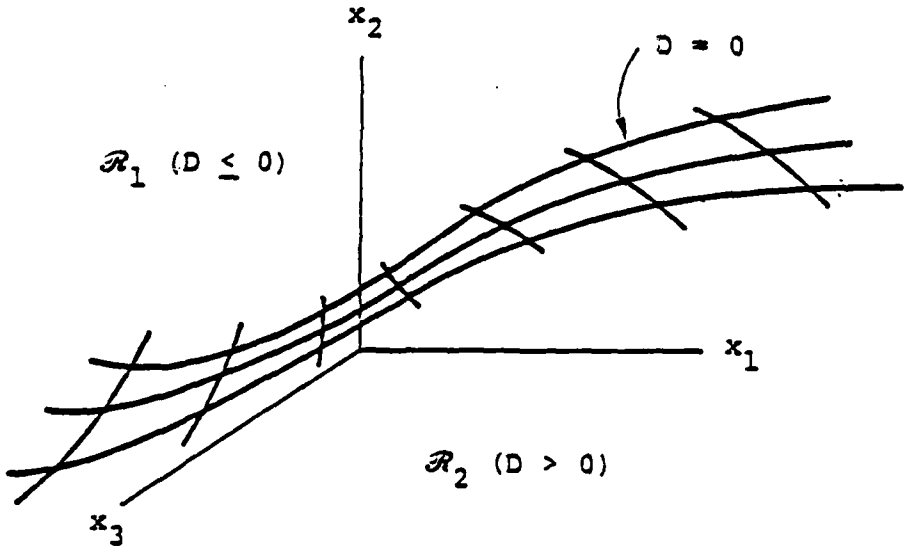


Figure 17. Schematic illustration of a decision surface in three-dimensional discriminant space.

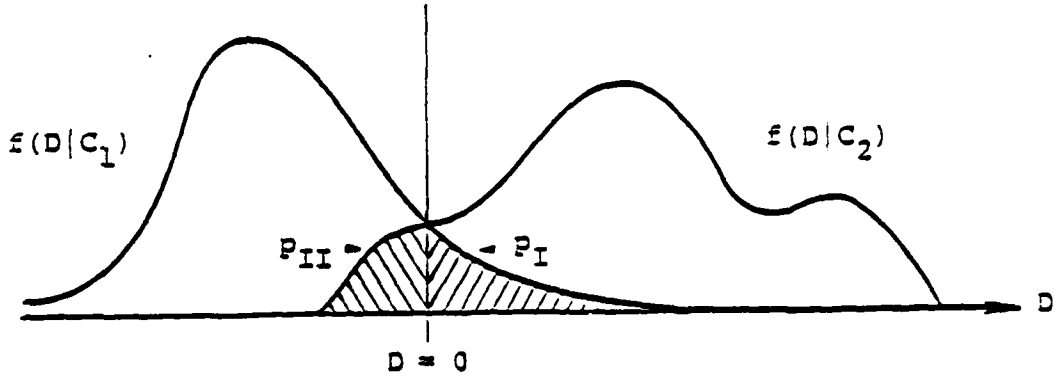


Figure 18. Illustration of the error probabilities of a decision rule in terms of the probability distributions of the decision function D.

$$\hat{C} = C_1 \quad \text{when} \quad \underline{x} \in \mathcal{R}_1$$
$$\hat{C} = C_2 \quad \text{when} \quad \underline{x} \in \mathcal{R}_2 \quad . \tag{3}$$

A classification $\hat{C}$ is useless without a measure of its accuracy. Misclassification (or error) probabilities, p_I and p_{II}, serve this purpose. We define p_I as the probability of assigning an event to Class 2 ($\hat{C} = C_2$) when it really came from class 1 ($\hat{C} = C_1$). Similarly for p_{II}. In terms of the probability distributions of $\underline{x}$ and D, the error probabilities are

$$P_I = \int_{\mathcal{R}_2} d^m x \; f(x|C_1) = \int_0^\infty dD \; f(D|C_1)$$
$$P_{II} = \int_{\mathcal{R}_1} d^m x \; f(\underline{x}|C_2) = \int_0^\infty dD \; f(D|C_2) \; . \tag{4}$$

Figure 18 illustrates these difinitions. We note that for a given function $D(\underline{x})$, the univariate distribution $f(D|C)$ is determined by the multivariate distribution $f(\underline{x}|C)$.

The decision approach involves deriving the decision function that minimizes the error probabilities in some sense. The Bayes criterion, for example, chooses $D(\underline{x})$ to minimize the expected "risk" of misclassification defined by

$$\rho = P(C_1) \; c_I \; p_I + P(C_2) \; c_{II} \; p_{II}, \tag{5}$$

where $P(C_1)$ and $P(C_2)$ are prior probabilities and c_I and c_{II} are assigned costs of misclassification. For $P(C_1) = P(C_2) = 1/2$, $c_I = c_{II}$, the decision function minimizing ρ becomes

$$D(\underline{x}) = \log f(\underline{x}|C_2) - \log f(\underline{x}|C_1) \quad . \tag{6}$$

Equations (1), (4) and (6) provide theoretical solutions to the discrimination problem when the distributions $f(\underline{x}|C_1)$ and $f(\underline{x}|C_2)$ are known. In the seismic discrimination problem, as in most scientific problems, the distributions are not known. A

variety of approaches have been developed to handle this situation, including nonparametric approaches which assume no knowledge of $f(\underline{x}|C)$ (e.g., nearest neighbor methods) and parametric approaches which require the estimation only of certain parameters of $f(\underline{x}|C)$, such as its mean and variance. We will not attempt to review the various methods here. The following sections describe a particular approach we have developed which deals realistically with the limited training sets available in the seismic discrimination problem.

3.2 FISHER LINEAR DISCRIMINANT

In designing a statistical method for seismic discrimination, we have kept the following considerations in mind:

- The method should require as few assumptions about $f(\underline{x}|C)$ as possible.
- Only a limited number of parameters of $f(\underline{x}|C)$ can be estimated accurately from the training sets.
- The method should provide realistic estimates of the error probabilities (p_I and p_{II}) associated with the decision function $D(\underline{x})$.
- The algorithm for obtaining $D(\underline{x})$, p_I and p_{II} (and posterior probabilities if they are desired) should be computationally efficient and suitable for automation.

Within these restrictions, the method should find the best decision function; i.e., the one with the smallest error probabilities.

A rather simple approach that can meet the above requirements is linear discrimination, which assumes a decision function of the form

$$D(\underline{x}) = \underline{a}^T\underline{x} + b \quad , \tag{7}$$

where b is a scalar and $\underline{a}$ is a vector containing M coefficients, or weights. Each coefficient in $\underline{a}$ multiplies one of the discriminants in $\underline{x}$. The decision surface $D(\underline{x}) = 0$ becomes a hyperplane in M-space. The objective in this approach is to find the $\underline{a}$ and b that minimize the error probabilities.

From Equation (6) it is apparent that a linear discriminant function does not provide an absolute minimum to p_I and p_{II} unless $f(\underline{x}|C_1)$ and $f(\underline{x}|C_2)$ have a particular functional form: the two distributions must be Gaussian with equal covariance matrices. However, the possible non-optimality of linear discriminants is of little consequence if there are insufficient training data to establish that the true distributions are significantly non-Gaussian, or have unequal covariances. Even then, the best linear discriminant might perform satisfactorily.

It is important to realize that with $f(\underline{x}|C)$ unknown, one cannot determine the error probabilities of any discriminant function exactly, but can only obtain estimates inferred from the training data sets. As a consequence of this, it may be impossible to distinguish the error probabilities of linear discriminants from those of nonlinear discriminants. If existing or future training sets prove adequate for making this distinction, straightforward extensions of our method to nonlinear functions can be made.

Because p_I and p_{II} must be estimated, the criterion that $D(\underline{x})$ minimize these probabilities does not lead to a direct algorithm for obtaining the optimal $\underline{a}$ and b. Therefore, we define a different criterion, one that has the effect of making the error probabilities small, and which leads to a direct algorithm for $\underline{a}$ and b. After $\underline{a}$ and b are obtained, p_I and p_{II} can be estimated to see how good the resulting $D(\underline{x})$ actually is.

The criterion we use to obtain the optimal linear $D(\underline{x})$ is given by the Fisher linear discriminant. The Fisher discriminant attempts to maximize the separation between the distributions $f(D|C_1)$ and $f(D|C_2)$. Separation is defined in terms of the means and variances of the distributions, as estimated from the training data. For a given $\underline{a}$ and b, the training sets produce sample values of D for each class, which we denote $D_{i(1)}$, $i = 1, .., N_1$, and $D_{i(2)}$, $i = 1, .., N_2$:

$$D_{i(k)} = \underline{a}^T \underline{x}_{i(k)} + b \qquad (8)$$

where k may be 1 or 2. We let $m_{D(k)}$ and $s^2_{D(k)}$, respectively, denote estimates of the mean and variance of $f(D|C_k)$ obtained from the samples $D_{i(k)}$. Then the Fisher discriminant satisfies

$$v^2 = \frac{(m_{D(2)} - m_{D(1)})^2}{s^2_{D(1)} + s^2_{D(2)}} = \text{maximum}$$

$$m_{D(2)} - m_{D(1)} = 1 \tag{9}$$

Figure 19 illustrates this measure of separation with a simple example.

To express v^2 in terms of $\underline{a}$, we let the vector $\underline{m}_{(k)}$ be an estimate for the mean of $f(\underline{x}|C_k)$ and the matrix $S_{(k)}$ be an estimate for the covariance matrix of $f(\underline{x}|C_k)$. For example,

$$\underline{m}_{(k)} = \frac{1}{N_k} \sum_{i=1} \underline{x}_{i(k)}$$

$$S_{(k)} = \frac{1}{N_k - 1} \sum_{i=1} (\underline{x}_{i(k)} - \underline{m}_k)(\underline{x}_{i(k)} - \underline{m}_k)^T \quad . \tag{10}$$

Fancier estimates can also be used. We then have

$$m_{D(k)} = \underline{a}^T \underline{m}_{(k)} + b$$

$$s^2_{D(k)} = \underline{a}^T S_{(k)} \underline{a} \tag{11}$$

so Equation (9) becomes (Gnanadesikan, 1977, p. 83; Young and Calvert, 1974, Equation (4.86))

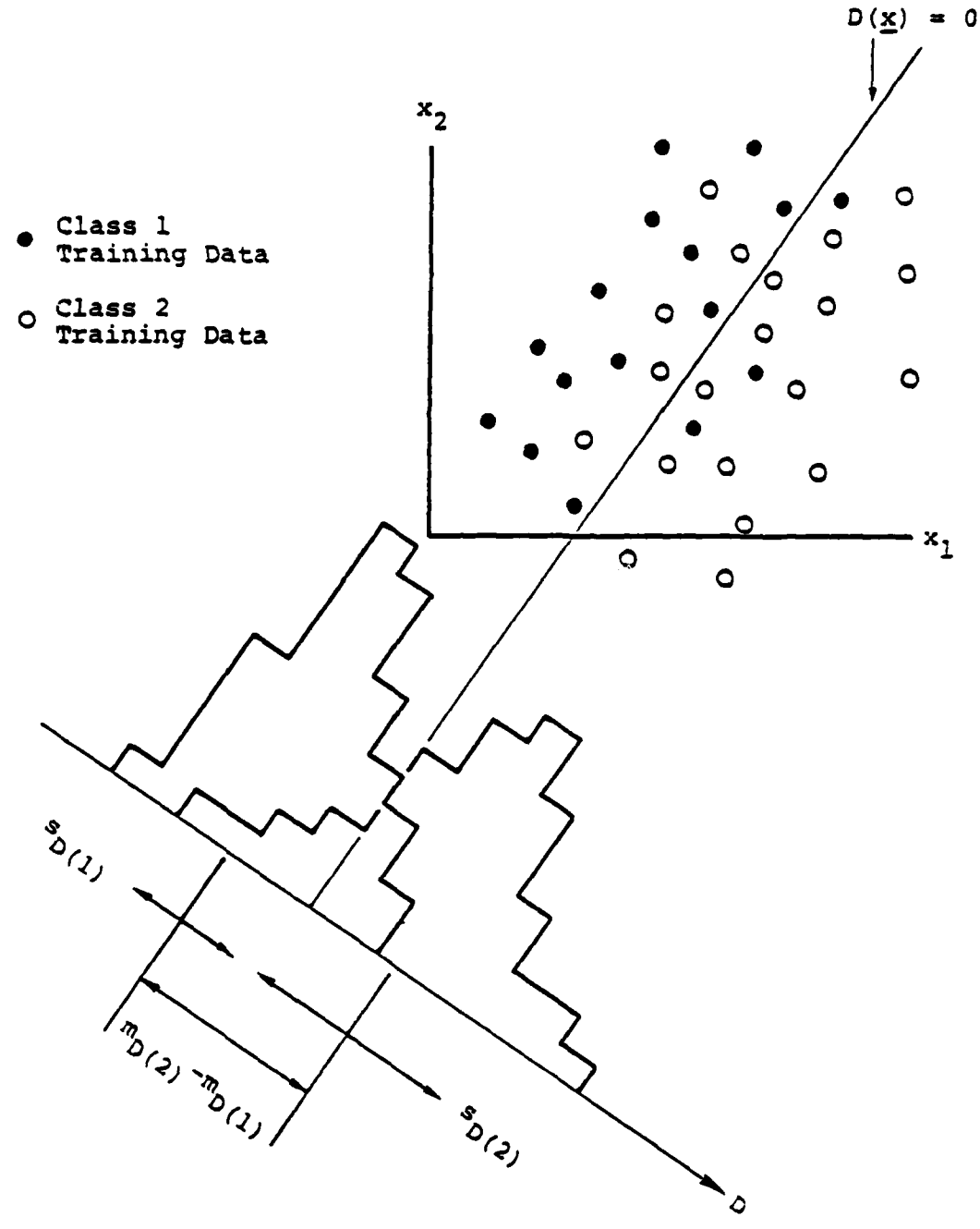


Figure 19. Illustration in two-dimensions of the separation of two classes of training data by a linear discriminant.

$$v^2 = \frac{(\underline{a}^T \Delta \underline{m})^2}{\underline{a}^T S \underline{a}} = \text{maximum}$$

$$\underline{a}^T \Delta \underline{m} = 1 \tag{12}$$

where

$$\Delta \underline{m} = \underline{m}_{(2)} - \underline{m}_{(1)}$$

$$S = S_{(1)} + S_{(2)} \quad . \tag{13}$$

The solution to Equation (12) is (Gnanadesikan, 1977, p. 84; Young and Calvert, Equation (4.93))

$$\underline{a} = \frac{S^{-1} \Delta \underline{m}}{\Delta \underline{m}^T S^{-1} \Delta \underline{m}} \quad . \tag{14}$$

It is clear that v^2 = maximum does not constrain b. The optimal choice of b is difficult to define without making assumptions about the functional form of f(D|C). The following value of b is a reasonable choice that tends to make p_I and p_{II} equal (Young and Calvert, equation (4.95))

$$b = \frac{-s_{D(1)} \underline{a}^T \underline{m}_{(2)} - s_{D(2)} \underline{a}^T \underline{m}_{(1)}}{s_{D(1)} + s_{D(2)}} \tag{15}$$

where $s_{D(k)}$ is given by Equation (11) and $\underline{a}$ by Equation (14). When $s_{D(1)}$ equals $s_{D(2)}$, b reduces to the constant which may be derived from the assumption that the two populations have equal covariance matrices (Anderson, 1958, Equation (6.4.5), p. 134).

3.3 FEATURE EXTRACTION BY DAMPING

$S_{(k)}$ (k = 1 or 2) is an M by M matrix which estimates the true covariance matrix of $f(\underline{x}|C_k)$. Its accuracy depends on the relative sizes of M (the dimension of $\underline{x}$) and N_k (the number of

samples of $\underline{x}$ in the training set). It is desirable to have $N_k >> M$. If N_k is too small, the inaccuracy of $S_{(k)}$ causes $S^{-1}_{(k)}$, and thus $\underline{a}$, to be unstable. In other words, $\underline{a}$ is weakly constrained by training sets with too few samples. When this occurs, $D(\underline{x})$ will perform well in classifying the training events but may perform poorly on new events. With a good algorithm for estimating p_I and p_{II} (e.g., jackknifing), this will be reflected in large estimates of these error rates.

The usual solution to this problem is feature extraction: a statistical procedure for transforming $\underline{x}$ to a vector $\underline{x}'$ of smaller dimension M'. (It is assumed that feature selection was done in creating $\underline{x}$ such that all the features in $\underline{x}$ are believed to be potentially good discriminants.) Given M', the desired number of new features, an ideal feature extractor would find the M' combinations of the original discriminants that could produce the best discriminant function $D(\underline{x}')$. It is very difficult, if not impossible, to do this since error probabilities cannot be estimated until after feature extraction has been performed.

A commonly applied method of feature extracation is the principal component method. In this method, the pooled covariance matrix of $\underline{x}$ (defined by Equations (10) and (13)) is decomposed into its eigenvectors and eigenvalues:

$$S = U\Lambda U^T \quad , \tag{16}$$

where the columns of U are the orthonormal eigenvectors of S and Λ is a diagonal matrix containing the eigenvalues of S. The principal component method takes as the new feature vector

$$\underline{x}' = U_L^T \underline{x} \quad , \tag{17}$$

where U_L is an M by M' matrix containing the eigenvectors associated with the M' largest eigenvalues. There is no guarantee that $\underline{x}'$ contains M' good discriminants since the projection of $\Delta\underline{m}$

onto U_L is not considered. (If $U_L^T \underline{\Delta m}$ were by misfortune zero, $\underline{x}'$ would be almost useless for discrimination.) Nevertheless, the principal component method has proved satisfactory in many applications (Tjostheim, 1981).

We have designed a variation on the principal component method which is more convenient and which promises to perform better because it takes $U^T \underline{\Delta m}$ into account. Feature extraction as such is not done. Instead, a damping term is added to S to form a new covariance estimate S(e):

$$\begin{aligned} S(e) &= S + eI \\ &= S_{(1)} + S_{(2)} + eI \end{aligned} \tag{18}$$

where e is a scalar damping parameter and I is the unit matrix. $S(e = 0) = S$, the undamped matrix used in the last subsection. We then obtain $\underline{a}$ by replacing S with S(e):

$$\underline{a}(e) = \frac{S(e)^{-1} \underline{\Delta m}}{\underline{\Delta m}^T S(e)^{-1} \underline{\Delta m}} \quad . \tag{19}$$

The damping of S is equivalent to adding e to each of its eigenvalues. Denoting the eigenvalues as λ_j and the associated eigenvectors (columns of U) as $\underline{u}_j$, $\underline{a}(e)$ takes the form

$$\underline{a}(e) \propto \sum_{j=1}^{M} \underline{u}_j \left(\frac{\underline{u}_j^T \underline{\Delta m}}{\lambda_j + e} \right) . \tag{20}$$

Thus e diminishes the contribution of $\underline{u}_j$ to $\underline{a}$ when λ_j is small, which is consistent with principal component feature extraction. However, $\underline{u}_j$ will still contribute to $\underline{a}$ if $\underline{u}_j^T \underline{\Delta m}$ is sufficiently large; i.e., if the means of the two event classes differ significantly in the $\underline{u}_j$ direction.

A refinement to both the principal component method and the damping method is achieved by using the correlation matrix R in the analysis instead of S. R is a scaled version of S:

$$R = W^{-1/2} \, SW^{-1/2} \tag{21}$$

where W is a diagonal matrix containing the diagonal elements of S ($W_{ij} = S_{ij} \, \delta_{ij}$, implying $R_{ii} = 1$). The eigenvalues of R are independent of the units chosen for $\underline{x}$; so the individual discriminants in $\underline{x}$ are normalized in a natural way. For the damping method, this refinement corresponds to redefining S(θ) as

$$S(\theta) = S + \theta W \quad . \tag{22}$$

It is not obvious how to determine the optimal value of θ; i.e., the value that results in the smallest error rates for $D(\underline{x})$. One could base a choice of θ on the expected uncertainties in S inferred under the hypothesis of a particular distribution $f(\underline{x}|C_k)$. A trail-and-error approach might be more effective, however; namely, one could compute $D(\underline{x})$ for several θ's and select the one yielding the smallest estimated error rates.

Examples of damped discriminant weights $\underline{a}(\theta)$ are shown in Figures 20 and 21. The data used in these examples are VFM magnitudes determined at two SRO stations: KAAO (Kabul, Afghanistan) in the first example (Figure 20) and CHTO (Chiang Mai, Thailand) in the second example. The training events are the AI events used in the Discrimination Experiment. The set of training events is not identical in the two examples, mainly because teleseismic events were excluded from the KAAO training set.

The feature vector $\underline{x}$ in each example has dimension 40 and contains $m_b(f)$ values at 40 frequencies between 0.4 and 5.0 Hz. From these 40 data, $m_b(f)$ at two frequencies (one high and one low) were selected as features for the Discrimination Experiment. The purpose of the examples shown here is to see what the statistical analysis determines as the optimal combination of the

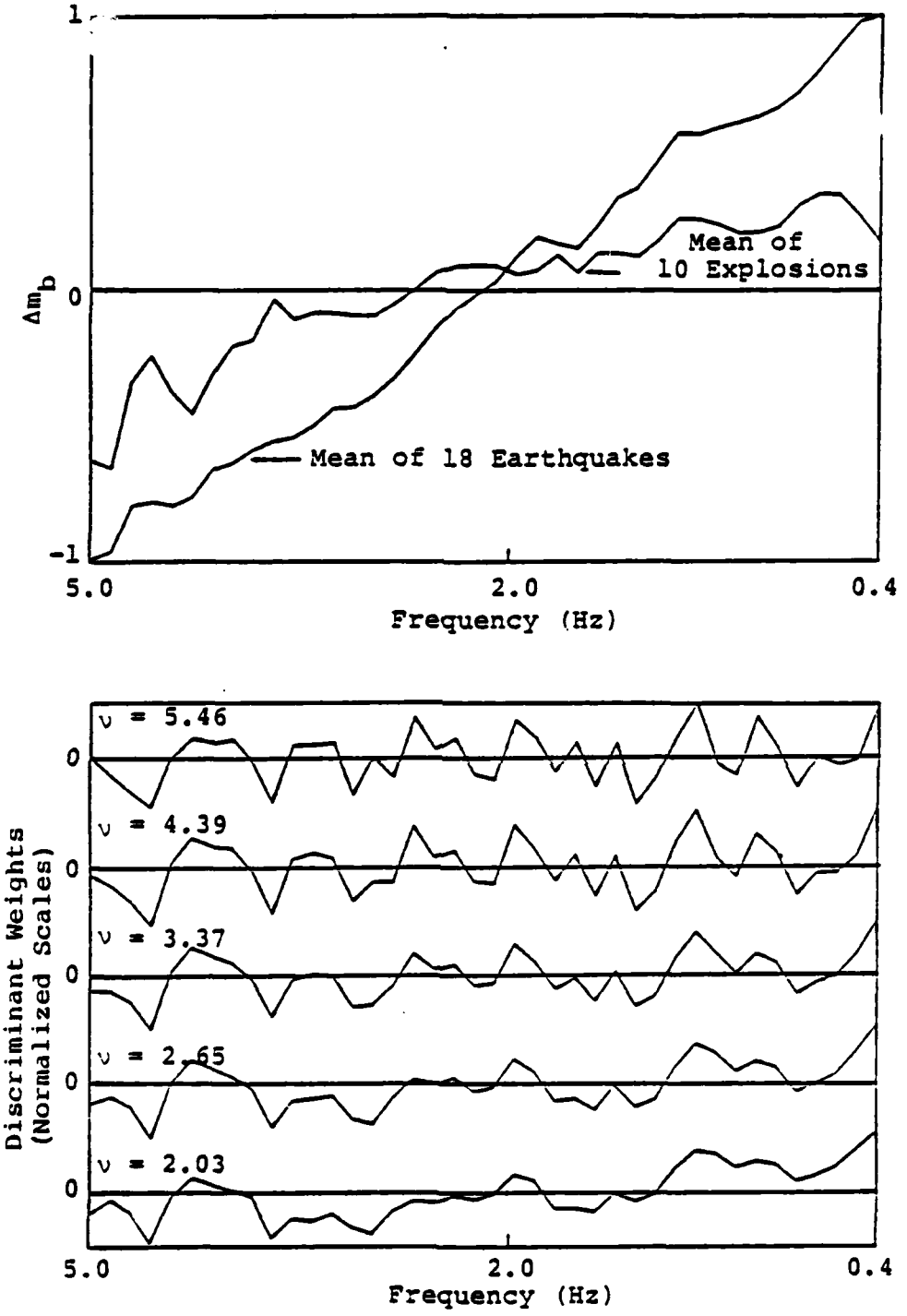


Figure 20. Training set means and five sets of feature weights determined from VFM data at station KAAO.

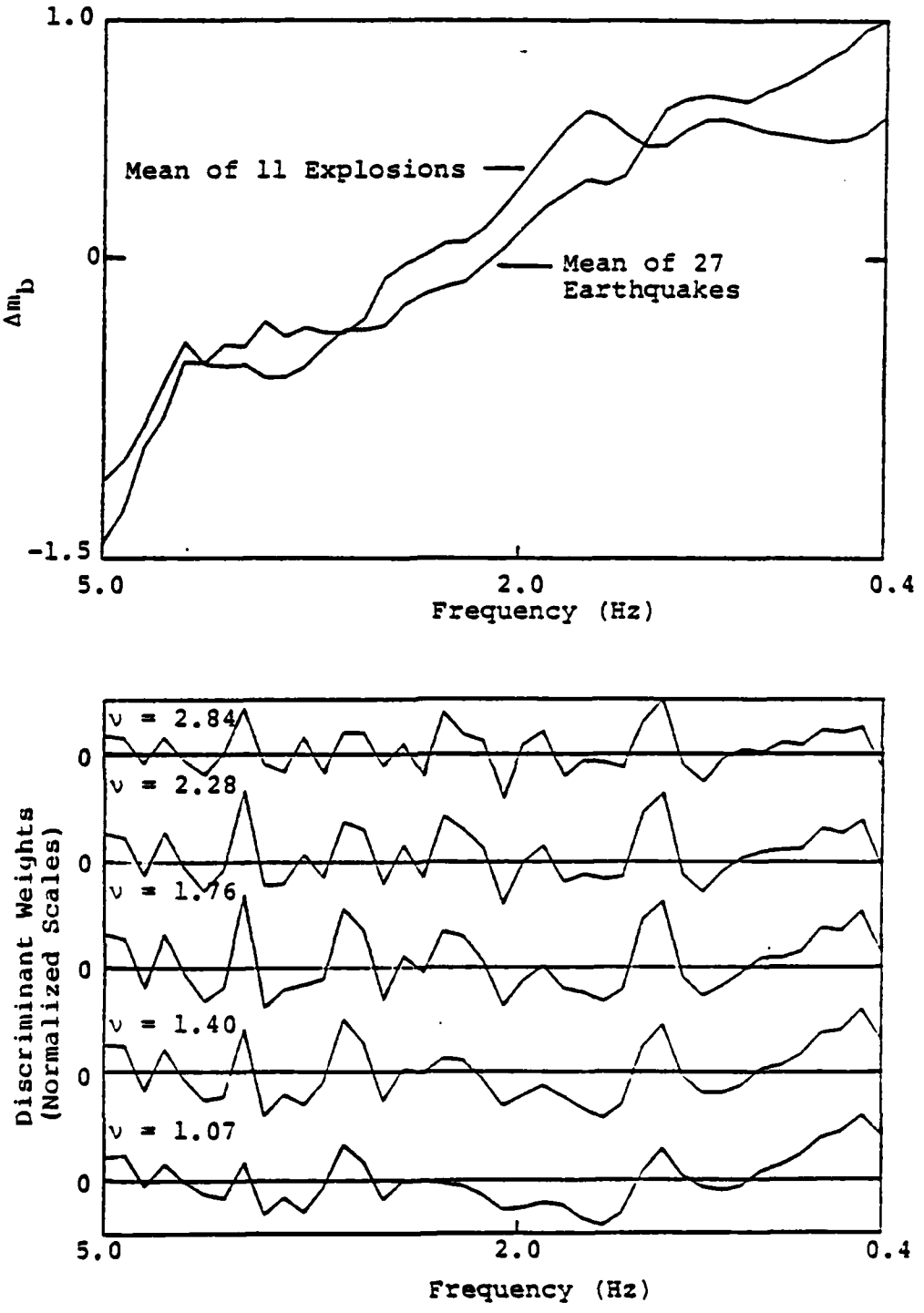


Figure 21. Training set means and five sets of feature weights determined from VFM data at station CHTO.

entire set of 40 m_b's. We point out that the Discrimination Experiment showed that the VFM discriminant from KAAO performed very well in classifying AI events, while the CHTO data performed poorly.

In each of Figures 20 and 21, the top frame shows the sample means of the explosion and earthquake training data ($\underline{m}_{(1)}$ and $\underline{m}_{(2)}$ from Section 3.2). The 40 components of each vector mean are plotted as a curve against a log frequency scale, and thus are displayed like a Fourier spectrum. We note that before doing the statistical analysis, the data were converted to a relative magnitude $\Delta m_{b(f)}$ by removing the average value of each feature vector (the average over frequency). Comparing Figures 20 and 21, we see that the separation between the explosion and earthquake means is much larger at KAAO than CHTO.

In each example, discriminant weights were computed for five values of the damping parameter θ, using the correlation damping scheme (Equation 22). The five sets of weights are plotted as a function of frequency in the bottom frame of each figure. The weights obtained with the smallest θ are plotted at the top of the frame and those with the largest θ (most damping) at the bottom. The zero line is drawn through each weight-versus-frequency curve, but the vertical scale for each curve is arbitrary and not shown.

Labeling each set of weights in Figures 20 and 21 is the value the weights give to the separation parameter ν (defined in Equation 12). This measures how well the discriminant plane $D(\underline{x}) = 0$ separates the two classes of training data. For our purposes here, $\nu \geq 2$ implies reasonably good separation. Comparing the two stations, we see that the KAAO data separate the event classes much better than the CHTO data.

The separation parameter decreases as the damping parameter increases. This does not mean, however, that the lowest damping provides the best discriminant function since good separation of training events does not imply low error rates on new events. In these examples where the number of training events is quite small compared to the number of features, the weights for small θ are probably unstable; so the heavily damped weights are likely to perform better on new events.

In both Figures 20 and 21, one can see that increasing the damping has the effect of smoothing the weights over frequency. Under-damped weights oscillate rapidly and attempt to use spurious wiggles in the $m_b(f)$ spectra as a basis for discrimination. The weights obtained with the most damping extract a very robust feature from the data, particularly in the KAAO example (Figure 20). They essentially subtract the average m_b below about 1 Hz from an average m_b over one or more high frequency bands - the actual bands varying from station to station. This is consistent with what we learned in the Discrimination Experiment.

3.4 JACKKNIFING TO OBTAIN ERROR ESTIMATES

When the probability distributions of $\underline{x}$ and $D(\underline{x})$ are unknown, the error rates p_I and p_{II} cannot be computed from Equation (4); they must be estimated empirically from the training data $\underline{x}_{i(k)}$. Estimates of p_I and p_{II} are easily obtained by counting the fraction of training events misclassified by $D(\underline{x})$; that is, the events that make $D_{i(1)} > 0$ or $D_{i(2)} < 0$ (see Equation 8). However, if D is derived from all of the training data, and thus optimized for these data, the error estimates may be very biased downward. This is particularly true when the dimensionality of $\underline{x}$ is high compared to the sample sizes.

A powerful method that removes much of this training set bias is the "leave-one-out," or "jackknife," method. Mosteller and Tukey (1977) provide a good discussion of jackknifing with illustrative examples. A simple application of jackknifing computes $\underline{a}$ and b N times ($N = N_1 + N_2$) leaving each training event out in turn. The discriminant function obtained each time is applied to the left-out event to obtain a sample, $D^*_{i(k)}$. The star distinguishes this from the sample $D_{i(k)}$ obtained with the complete discriminant function. The idea behind jackknifing is that the $D^*_{i(k)}$ are a more likely set of samples of $f(D|C_k)$ than are the $D_{i(k)}$. The number of $D^*_{i(k)}$ having the wrong sign thus produce less biased estimates of p_I and p_{II}. We

denote the jackknifed estimates as p_I^* and p_{II}^*. The variances of p_I^* and p_{II}^* are a function of the training sample sizes, being roughly inversely proportional to N_1 and N_2. More accurate error estimates might be obtained with more elaborate jackknifing procedures that yield more than N samples of D* (e.g., leave-two-out).

The jackknifed samples $D_{i(k)}^*$ can be used to estimate the complete probability distributions $f(D|C_1)$ and $f(D|C_2)$. An estimate of $f(D|C_k)$ may be obtained by fitting a smooth curve to the cumulative histograms of the $D_{i(k)}^*$, or by Parzen''s approximation. From estimates of the distributions, one can derive estimates of the Bayes posterior probabilities using D in place of $\underline{x}$ in Equation (1).

A drawback of the jackknife method is the considerable amount of computation it involves. Fortunately, the computations required by the Fisher discriminant are rather modest compared to many alternate approaches; so compuational considerations might not matter if the sample sizes and dimensionality are not too large. In addition, we have devised efficient algorithms for finding inverses of the perturbed covariance matrices that occur in the leave-one-out method. These algorithms would reduce the jackknifing computation by a factor of order N. The algorithms do not seem to be applicable to the damping scheme involving the correlation matrix (Equation 22). They are, however, applicable to the covariance damping scheme (Equation 18).

3.5 NONLINEAR DISCRIMINANTS

If we discover that the damped Fisher discriminant performs unsatisfactorily for automatic seismic discrimination, a particular nonlinear discriminant can be implemented with only modest modification of the algorithm outlined above. The decision function is quadratic in $\underline{x}$:

$$D(\underline{x}) = (\underline{x} - \underline{m}_{(1)})^T (S_{(1)} + \theta_1 I)^{-1} (\underline{x} - \underline{m}_{(1)})$$

$$- (\underline{x} - \underline{m}_{(2)})^T (S_{(2)} + \theta_2 I)^{-1} (\underline{x} - \underline{m}_{(2)}) \qquad (23)$$

$$+ \log [\det (S_{(1)} + \theta_1 I)/\det (S_{(2)} + \theta_2 I)] \ .$$

This is the Bayes decision function (Equation 6) implied by Gaussian $f(\underline{x}|C_1)$ and $f(\underline{x}|C_2)$, but with sample means and damped sample covariances substituted for the true means and covariances of the distributions. Even though we did not optimize any free coefficients (like $\underline{a}$) to derive the quadratic discriminant, it turns out that the Fisher discriminant is a special case of Equation (23). When $S_{(1)} = S_{(2)}$, $\theta_1 = \theta_2$, $D(\underline{x})$ reduces to the Fisher discriminant, but with a different value of b from that in Equation (15). In this sense, $D(\underline{x})$ in Equation (23) might be considered more optimal than the Fisher discriminant. However, this is not necessarily the case since only finite training sets are available for estimating the covariance matrices.

The algorithm for implementing this discriminant would not differ very much from the Fisher discriminant algorithm. The same sample means and covariance matrices are involved, and are just combined differently to obtain $D(\underline{x})$. The jackknifing procedure would proceed in the same way, including the shortcut algorithm we mentioned for inverting perturbed covariance matrices, if it is needed. The damping parameters θ_1 and θ_2 in Equation (23) stabilize $D(\underline{x})$ in the same way that θ stabilizes the Fisher discriminant. Like θ, they can be optimized by trial and error or selected on theoretical grounds.

Finally, we mention a variation on the linear discrimination approach which can be used to optimize, in a limited way, general nonlinear discriminants. The procedure, described by Young and Calvert (1974), is to augment the feature vector $\underline{x}$ with nonlinear functions of its original elements. An example illustrates the

basic idea. Let $\underline{x}'$ be an augmented feature vector containing squares and cross products of the elements of $\underline{x}$:

$$\underline{x}' = (x_1, x_2, .., x_M, x_1^2, x_1x_2, ..., x_M^2) \quad . \tag{24}$$

Then the linear discriminant

$$D(\underline{x}') = \underline{a}^T\underline{x}' + b \tag{25}$$

becomes a general quadratic function of the original $\underline{x}$. The Fisher linear discriminant algorithm applied to $\underline{x}'$ would, in effect, optimize a quadratic decision function. The danger in this approach, of course, is that augmentation increases the dimension of the feature vector which might cause stability problems.

Given limited training sets, it may be more beneficial to treat nonlinear functions of the data in the feature selection phase. If a nonlinear function of a datum makes its distributions more Gaussian-like (e.g., the z-statistic), its use as a feature in $\underline{x}$ will improve the performance of both the Fisher discriminant and the quadratic discriminant function in Equation (23).

3.6 REGIONALIZATION OF DISCRIMINANTS

Statistical discrimination methods assume that the training data within each event class are identically distributed as $f(\underline{x}|C_k)$ ($k = 1$ or 2), and that a new feature vector $\underline{x}$ to be classified has one of these two distributions. The error probabilities p_I and p_{II} reflect an average performance on events whose data have these distributions. These assumptions bear on two important aspects of the seismic discrimination problem.

First, if events from different source regions are analyzed together, then the method must treat the regional variations in the discriminants due to geology as a random process that disperses $f(\underline{x}|C_k)$. This always degrades average error rates and may, in fact, be a bad model. Thus, it is clearly necessary to divide the

total data set into subsets appropriate to particular source regions. In this case, each subset is treated independently using the machinery discussed in the previous sections. Our approach is, therefore, to set up a regionalized classification for events based on geophysical province and to use real or theoretically produced event training sets from such regions as independent populations. For example, we would classify all trench located earthquakes as a separate population of events. Similarly, we would treat events occurring within plate interiors as a separate population, and events from rift zones or ridges as a third population, and so on. With this regional subsetting procedure, we would anticipate far less dispersion in the individual populations than would be the case if all events were classed together in one population, and, consequently, much more meaningful estimates of error probabilities and more precise and accurate event classifications.

3.7 MEASUREMENT ERROR

The second feature of the actual data that affects our use of the previously discussed multivariate analysis is the errors and uncertainties in the observations. These errors are dependent on the receiver network distribution relative to the source location, the noise levels at the receivers and the event magnitude itself. Estimates of these errors and uncertainties can be made; and we have, for example, taken great care in obtaining noise related uncertainty estimates in our automated measurement of discrimination variables. The errors are not uniform in size with respect to event magnitude and location, however, and this fact requires that we either separate the populations into magnitude ranges and regions where the data measurement uncertainties are nearly uniform, or include the nonuniform data uncertainties in the multivariate discrimination procedures from the beginning. In the former case, the method outlined in the previous sections can be applied directly. For the latter treatment, where nonuniform error estimates are to be treated directly, the procedure formulated earlier would need to be generalized.

4. APPLICATION OF THE FISHER DISCRIMINANT TO SPECTRAL RATIO (VFM) MEASUREMENTS

In Chapter 3, a prescription for classifying single seismogram feature vectors was developed. The prescription was very simple, consisting of simply forming the dot product of the feature vector with a set of weights and then adding a constant, b, to yield the scalar discriminant which we call d. The weight vector, $\underline{a}$, and the constant, b, are obtained from analysis of training data which, of course, must contain samples from the two populations. Except for a scale factor, the weight vector, $\underline{a}$, is just that first derived by Fisher and is found by multiplying the inverse of the pooled variance-covariance matrix by the vector difference in the means of the two populations (see Equation 14). For the constant, b, we have followed the suggestion of Young and Calvert (see Equation 15) which, it may be shown, reduces to the classical Fisher result in the case where the covariance matrices of the earthquake population and explosion population are equal. In this chapter, we discuss the results obtained when this rule is applied to the variable frequency magnitude features calculated for some of the Area of Interest seismograms. Only three stations (KAAO, RKON, and ILPA) are mentioned, the purpose being to describe the method by actual illustration, in order to exemplify the procedures whereby a large set of feature weights have been obtained for incorporation into the automatic discrimination package.

It may be recalled that we pointed out several theoretical weaknesses to the application of the Fisher linear discriminant for earthquake explosion discrimination. One possible weakness is that the Fisher linear discriminant does not necessary minimize the P_I and P_{II} misclassification probabilities. Although this discriminant does minimize those probabilities in the case where the two populations have multivariate Gaussian probability density functions with equal covariance matrices, this is almost surely not the case for earthquake and explosion seismograms. Secondly, the constant, b, which essentially defines the decision rule is open to question,

for there seems to be little theoretical guidance on how the decision rule should be obtained for data such as we deal with. Finally, there is the problem of attaching confidence limits to parameters estimated from the training data. There exist, of course, multidimensional equivalents of the t statistic of the χ^2 statistic, etc., but, in view of the demonstrable difficulties in supporting the Gaussian assumption about our data, we are wary of applying the ordinary tests of significance to the quantities which we estimate from the training data. Not only are the sample sizes pitifully small, but also the training data are known to have inherent biases in such respects as propagation path, magnitude range, etc. Thus, any tests of significance based on the Gaussian assumption would be highly suspect.

To examine these sorts of questions, which are really questions about the robustness of the statistical methods which have been used, the analysis technique known as jackknifing (or the leave-one-out method) has been applied in conjunction with the work on the Fisher linear discriminant. The idea behind jackknifing is trivially simple. Given training data from each of the two populations, one simply pretends that one of the datums (feature vectors) was not available for the analysis. Under the fiction that the diminished training data is the only data available, the discrimination function, $\underline{a}$, (weights vector) and decision constant, b, are evaluated using the methods described in Chapter 3. This rule is then applied to the single datum which was left out of the analysis. This results, for the ignored feature vector, a classification scalar, d*. This single scalar is saved and tagged with the identifier of the datum which was left out. The previously ignored datum is then put back into the training set, another datum is dropped from consideration, and the analysis repeated until one has cycled through the whole set of training data. If there are n data vectors available, this then requires n evaluations of the Fisher linear discriminant.

A jackknife analysis has several attractive features. Probably the principal one is that it leads to misclassification probabilities which more clearly reflect the operational environment in which seismic discrimination must be carried out. It mimics the situation where the one freezes the training data, derives and algorithm, and then applies the algorithm to new measurements. We find, in fact, that misclassification probabilities obtained through the jackknifing method are more pessimistic, (that is, discrimination more error-prone) than are those which one would infer if all the training data were processed at once. A further desirable feature is that plots of the scalar discriminant d* obtained by the jackknifing graphically illustrate the tradeoff which is obtained when the decision rule, that is, the constant, b, is altered. Jackknifing is also a clear way of demonstrating the existence of outliers in the data, anomalous seismograms which may unduly bias the results. This is a phenomenon sometimes referred to in univariate statistics as a leverage point. In this respect, jackknifing fits neatly into the philosophical approach of Gnanadesikan (1977, p. 196), "The main function of statistical data analysis is to extricate and explicate the informational content of a body of data."

A major objection to jackknifing in the past seems to have been the computational expense entailed by the multiplicity of statistical calculations. This has been greatly exaggerated. For example, 60 feature vectors, each of dimension 40, can be processed in a minute or so on the UNIVAC 1100/81 computer, and the bulk of the AI VFM data set was processed in this project in a few man weeks -- and most of this time was taken up with data base preparation and not the statistical calculations themselves. Compared to the effort required first to organize the waveform data base and to edit it in preparation for feature extraction, the linear discriminant analysis and the jackknife calculation is short and simple. It presents the analyst with a wealth of information from the training data -- more data than can easily be absorbed and synthesized.

4.1 DESCRIPTION OF PROGRAM MVSD

A block diagram of the program which was used to estimate the station dependent feature weights, and to estimate the misclassification probabilities by jackknifing is shown in Figure 22. Because the Area of Interest variable frequency magnitude data were most accessible to us, these were the data used in the analyses. The method, of course, is applicable to a feature vector of arbitrary dimension, and further work in automatic discrimination will entail the expansion of this calculation to include other discriminants.

The first stage in the analysis is a data preparation step (see Figure 23). This is the only step which needs to be altered in order to include other discriminant data sets. The procedure begins by reading a set of execution parameters. This step continues in the second box by reading the set of feature vectors which are available for a single seismic station. This, of course, must include a selection of earthquake, as well as explosion, seismogram readings. In our application, we have worked with the 40 variable frequency magnitudes described by Savino, et al. (1981a) as contained on the widely distributed data tape. Generally, only a subset of all the available data is processed at a time. The data is typically partitioned, for example, into different magnitude ranges, or into different source regions, or into different distance ranges. The sorted list of feature vectors is then rearranged. For the case of the VFM discriminant where no magnitude was occasionally reported for some frequencies at which there were holes in the spectrum, any missing data were linearly interpolated.

After the raw data is acquired, the statistical analysis of the entire training set is performed. This breaks down into two major functions — the mean and covariance matrix calculations (see Figure 24) and the linear discriminant estimation (see Figure 25). The statistical procedure begins by computing new reference values based on selected events and then subtracts out from the VFM magnitudes for each event, the mean body wave magnitude. Thus, if one was to average the spectral magnitudes for each separate event,

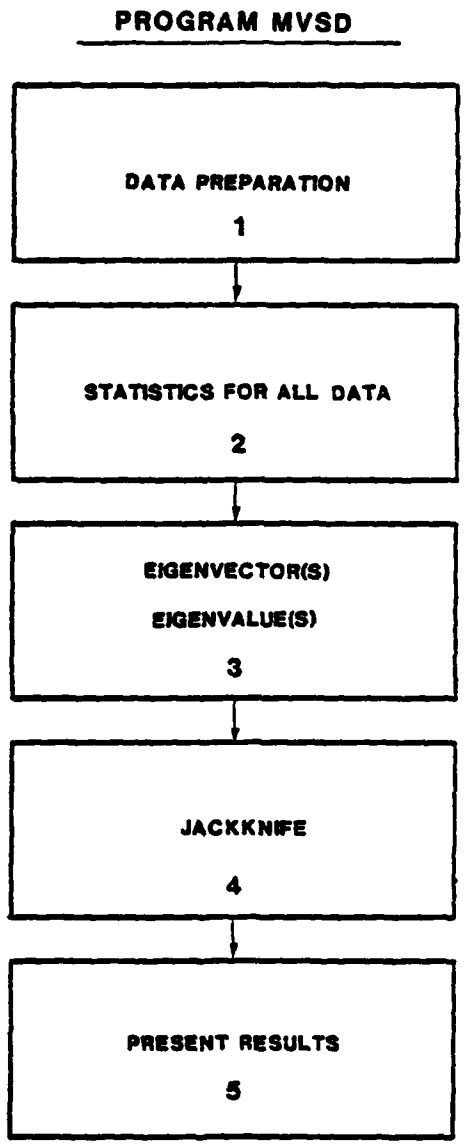


Figure 22. There are five steps to the procedure which estimates feature weights and misclassification probabilities.

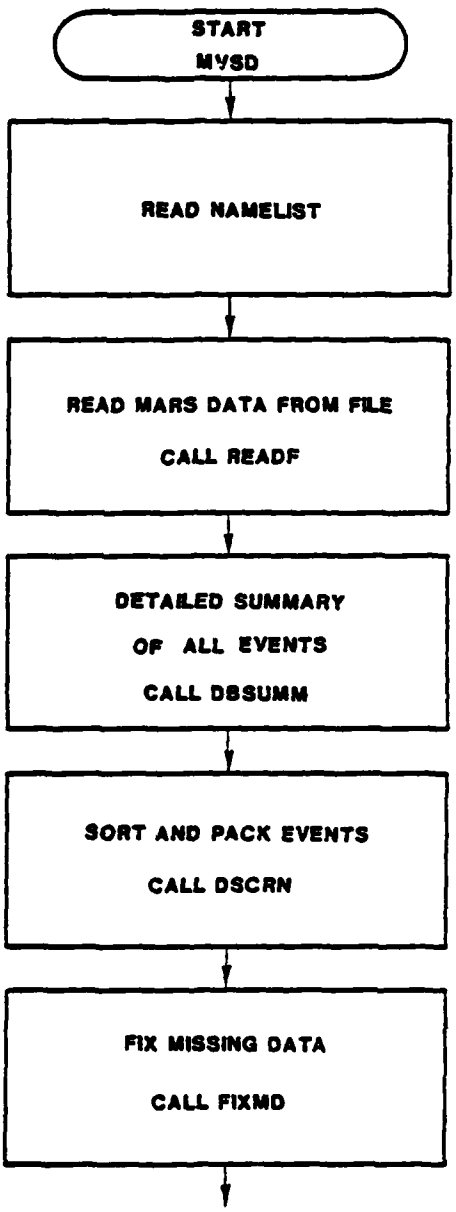


Figure 23. Data preparation is step one in program MVSD.

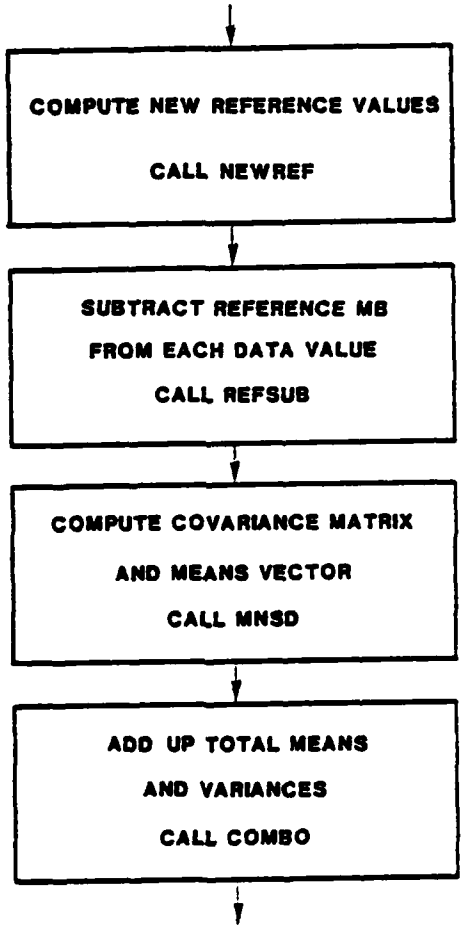


Figure 24. Statistical calculation on all training data for both event classes is step two.

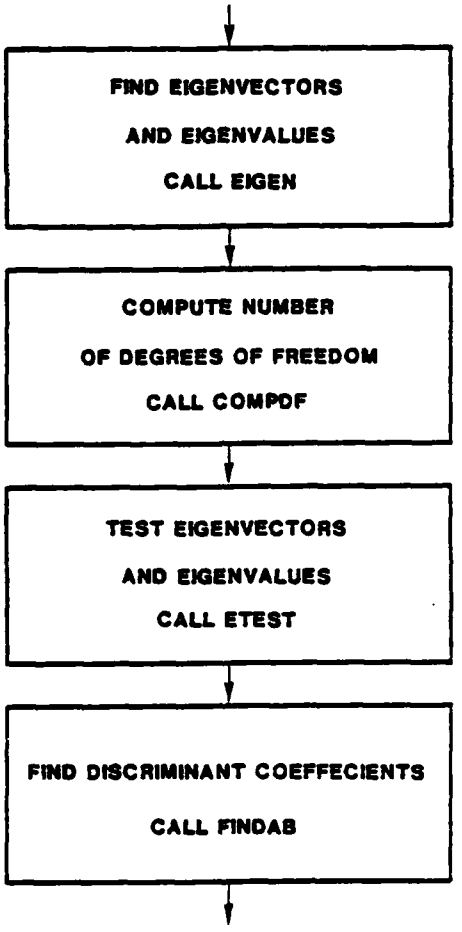


Figure 25. Eigenvector decomposition of the two covariance matrices is used to stabilize the feature weights.

the average of the modified magnitudes would be zero. Taking all feature vectors together, the mean vectors and the covariance matrices are found in the usual way (see Equation 10). From these, the total mean and variances are found, and, finally, the combined correlation matrix and the standard deviation.

Having found mean vectors and covariance matrices for the set of training data in each population, the weight vector, $\underline{a}$, which best separates the two populations, and the constant, b, which forms the decision rule, are then calculated (Figure 25). To do this, we first compute the eigenvalues and the eigenvectors of the combined covariance matrix using singluar value decomposition. The number of degrees of freedom of the linear system is then found, and the matrix of the eigenvectors is tested for orthonormality. Further tests are applied to the norm and to the trace of the matrix, for these tests are required to ensure the matrix is not singular. Then, the discrimination coefficents (feature weights) are found, and at this stage the parameter e (the damping parameter) is added in order to suppress the small eigenvalues of the variance covariance matrix. From the weights vector and the decision constant, b, the mean and the standard deviation of the scalar discriminant, d, for the entire data set are found.

The program next enters the inner jackknifing loop (see Figure 26) which is essentially a repeat of step 2, the statistical calculation, and step 3, the singular value decomposition and discriminant calculation, but for the diminished set of training data which results when each measurement vector in turn is left out of the analysis. Finally, the results are summarized, stored on a printed file, and partially printed on the line printer (Figure 27).

4.2 KAAO RESULTS

The purpose of this, and the two succeeding sections is to illustrate the practical application of the Fisher discriminant and jackknifing to the variable frequency magnitude data from the Area of Interest experiment. Although practically the entire VFM data

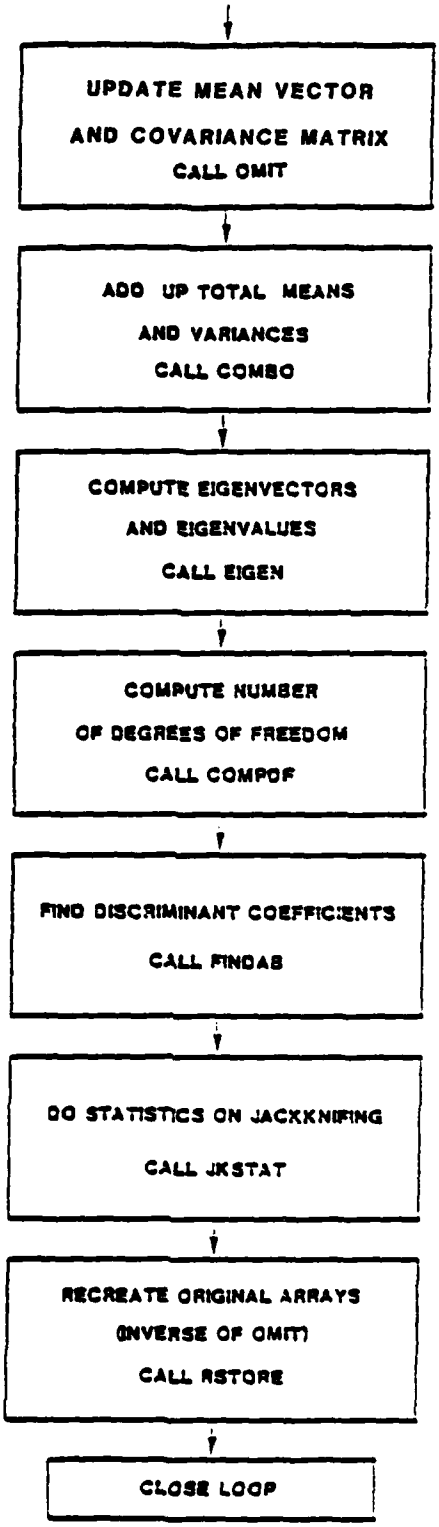


Figure 26. Jackknifing (the leave-one-out method) is used to estimate misclassification probabilities for the training data.

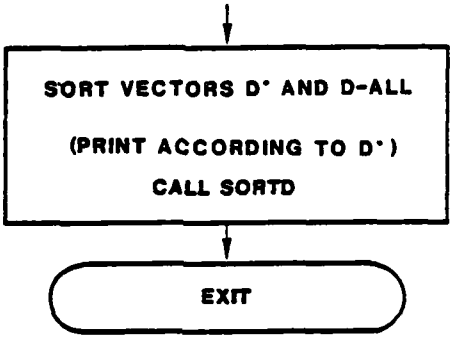


Figure 27. Jackknife results are printed and weights vectors saved to use in the Automatic Discrimination Program at SDAC.

set has been analyzed, we have extracted from those results what seem to be three representative examples -- results obtained for Kabul (KAAO), Redlake, Ontario (RKON), and the Iranian array (ILPA). Not only are the linear discriminant results and the jackknife results presented, but we also tie these results to those obtained by application of the bivariate discrimination procedure used previously by Savino. The comparison with Savino's earlier results shows that the importance of the programming error discovered by Rivers (1981) has been somewhat overestimated. Rivers found that, for small magnitude events, there were demonstrable mistakes in the extrapolation of the spectral magnitude to low frequencies. What we find in the analysis reported here, is that when no smoothing and extrapolation is done, that is, when all VFM magnitudes are taken exactly as measured, there is the same clear separation of the explosion and earthquake populations. This re-analysis of the VFM data set does not, of course, address the controversial issue of the physical basis for this discriminant. That is, whether it reflects a bias in the data set, whether it is a consequence of attenuation along the various propagation paths, or whether it truly arises right at the source.

The result of pilot calculations for the KAAO VFM data were presented in Chapter 3 (see Figure 20). The principal purpose behind those calculations was to explore the range of damping parameter, θ, which provides acceptable tradeoff between resolution and variance. It is a general observation that covariance matrices calculated for highly correlated random variables have a rank much less than the dimension (in this case, 40) of the matrix so that when its inverse is found, in order to estimate the set of weights, $\underline{a}$ (see Equation 14), the 40 separate weight factors may be highly erratic. The principal components analysis of the Kabul variance-covariance matrix, performed by singular value decomposition, shows that between 10 percent and 50 percent of the eigenvectors contained most of the variance in the data. On the basis of this observation, the entire VFM data set was processed using three values for the damping parameter. These are referenced in the subsequent figures

by the parameter K which takes the values 1, 10, or 30 in the three cases. Since the damping parameter is a rather abstract notion, we also tabulate, in each of the figures, the numbers of degrees of freedom (NDF) which apply in the three cases. These are roughly equivalent to the number of free parameters out of the 40 possible which were retained in the inversion of the variance covariance matrix. The value of NDF, when rounded to the nearest integer, is roughly equivalent to the dimension of the hyperplane which cleaves the data into the two populations. For bivariate discrimination, NDF would equal 2.

Figure 28 presents the results obtained from the linear discriminant analysis of 29 events (see Table 1) in the Kabul VFM data set. There are 10 explosions (Type -1) and 19 earthquakes (Type 1). As was mentioned in the discussion of Figure 24, the first step in the analysis is to subtract the mean magnitude from each 40 element VFM data vector. This gives 29 relative VFM vectors. Then, at each of the 40 frequencies, the sample mean and the sample standard deviation is calculated for each of the two populations. This results in an "average" earthquake spectrum and an "average" explosion spectrum, with accompanying relative m_b limits for the two classes which encompass 95 percent of the observations. Plots of the upper and lower limits, drawn around the average, are shown at the bottom of Figure 28. This figure indicates that the explosion data are relatively richer in high frequencies than are the earthquake data. We recall that if each of these bands is averaged across frequency, each would have a relative m_b of zero.

This initial step of subtracting the mean magnitude from each spectrum is controversial. The result may be expected to depend on the range of frequencies spanned. It takes no account of the frequency dependent signal-to-noise ratio, a particularly severe problem for weak events. Most importantly, it does not account for the fact that the shape of both earthquake and explosion spectrums vary with the moment (or size) of the event. Graphically, one can

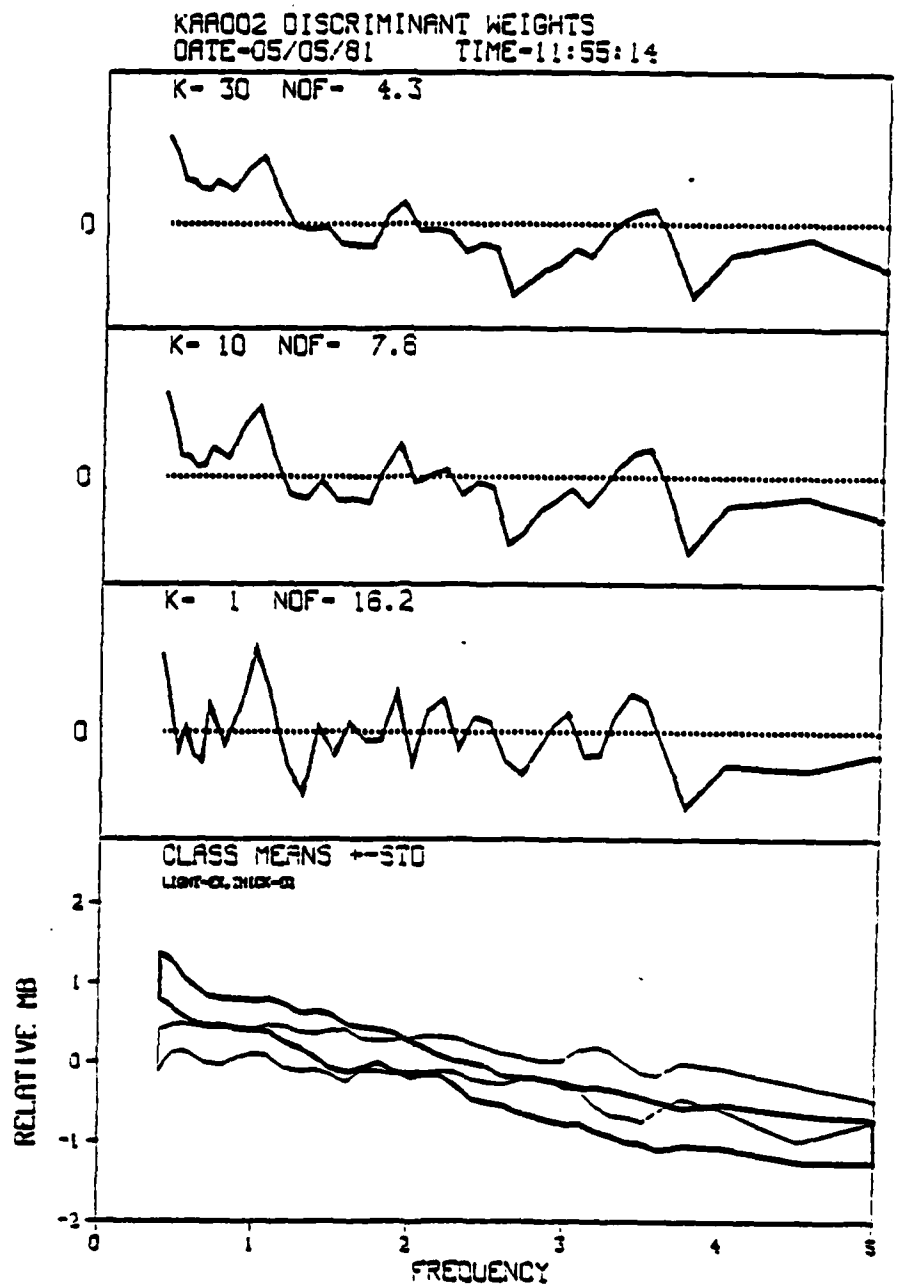


Figure 28. A plot of the feature weights vectors for the VFM discriminant at KAAO (three top panels) shows that as the number of degrees of freedom (NDF) increases, the weights become more "noisy". The trend in the weights clearly reflects the differences in the mean spectra for the two classes of events (bottom).

TABLE 1

KAAO SORTED JACKKNIFE SUMMARY

INDEX	EVENT	DELTA	MBREF	TYPE	CLASS	D-ALL	ORIGINAL ERRORS	DSTAR	JACKKNIFE ERRORS
1	14	16.6	3.70	-1	1	-.58		-.65	
2	267	17.1	5.02	-1	1	-.56		-.59	
3	189	16.7	4.10	-1	1	-.46		-.44	
4	17	16.7	4.28	-1	1	-.44		-.38	
5	20	17.2	5.41	-1	1	-.40		-.37	
6	271	16.8	4.95	-1	1	-.41		-.37	
7	266	16.8	4.26	-1	1	-.27		-.16	
8	81	17.8	4.87	-1	1	-.35		-.10	
9	272	17.7	3.49	1	2	.17		-.07	2
10	53	16.9	4.89	-1	1	-.18		.03	1
11	159	15.9	3.25	1	2	.34		.16	
12	22	20.5	4.59	-1	1	-.10		.19	1
13	152	23.5	3.72	1	2	.35		.29	
14	34	16.2	3.39	1	2	.42		.38	
15	195	18.2	4.51	1	2	.45		.38	
16	73	15.4	4.45	1	2	.53		.41	
17	29	19.40	3.58	1	2	.48		.41	
18	80	18.8	3.31	1	2	.50		.43	
19	46	20.8	3.10	1	2	.52		.45	
20	27	22.2	4.80	1	2	.55		.48	
21	163	20.4	3.50	1	2	.52		.48	

TABLE 1 (continued)

INDEX	EVENT	DELTA	MBREF	TYPE	CLASS	D-ALL	ORIGINAL ERRORS	DSTAR	JACKKNIFE ERRORS
22	166	23.7	4.72	1	2	.66		.60	
23	39	16.2	3.41	1	2	.66		.65	
24	31	16.3	3.46	1	2	.74		.75	
25	32	16.3	3.55	1	2	.85		.90	
26	68	15.8	3.19	1	2	.91		.98	
27	151	20.9	3.76	1	2	.90		.99	
28	35	16.2	3.92	1	2	.94		1.01	
29	162	20.5	4.02	1	2	1.17		1.35	

The summary listing of the jackknife results for the VFM data at KAAO shows that three events were misclassified according to this procedure, but when the data are taken all together, there is a perfect separation of the two populations.

understand the mean magnitude scaling by imagining a bivariate plot of spectral amplitudes, such as Figure 36 shown later. Removing the mean from each spectrum is tantamount to assuming that the data are clustered in ellipsoids with the major axis of each population meeting the coordinate axes at a 45 degree angle.

For three values of the damping parameter K, the linear discriminant analysis produces the three weights vectors shown in the top three panels of Figure 28. At the top of this figure, we see that with the damping parameter of 30, there are approximately 4.3 degrees of freedom for this data set whereas, when the damping parameter is 1, there are over 16 equivalent degrees of freedom. It is observed, as was pointed out in the discussion of the pilot calculation in Chapter 3, that when the number of degrees of freedom increases, the progression of weights becomes more and more erratic with frequency. Although the pattern of weight amplitudes for the K = 1 result is difficult to perceive, for K = 10 and K = 30, it is obvious that the set of weights tends to be positive at low frequencies and negative at high frequencies. This is, of course, just the observation upon which Savino founded his bivariate discrimination criterion.

For each of the three choices of damping parameter, the jackknife calculation was performed for all 29 events in the data set. The result of this calculation is presented in Figure 29. Recall that jackknifing consists of deleting, one at a time, each of the events from the data set and recalculating the best linear discriminant; that is, the set of weight vectors which best separates the residual members in the two populations. Thus, for each choice of the damping parameter, there were 29 sets of weight vectors calculated, but we have not displayed those here. What we show instead is the result of taking those 29 weight vectors and then using them to classify the single datum which was left out of the linear discriminant analysis. The results are presented as univariate plots of the scalar discriminant d* for all 29 events. In Figure 29 we represent the explosion events by open circles and

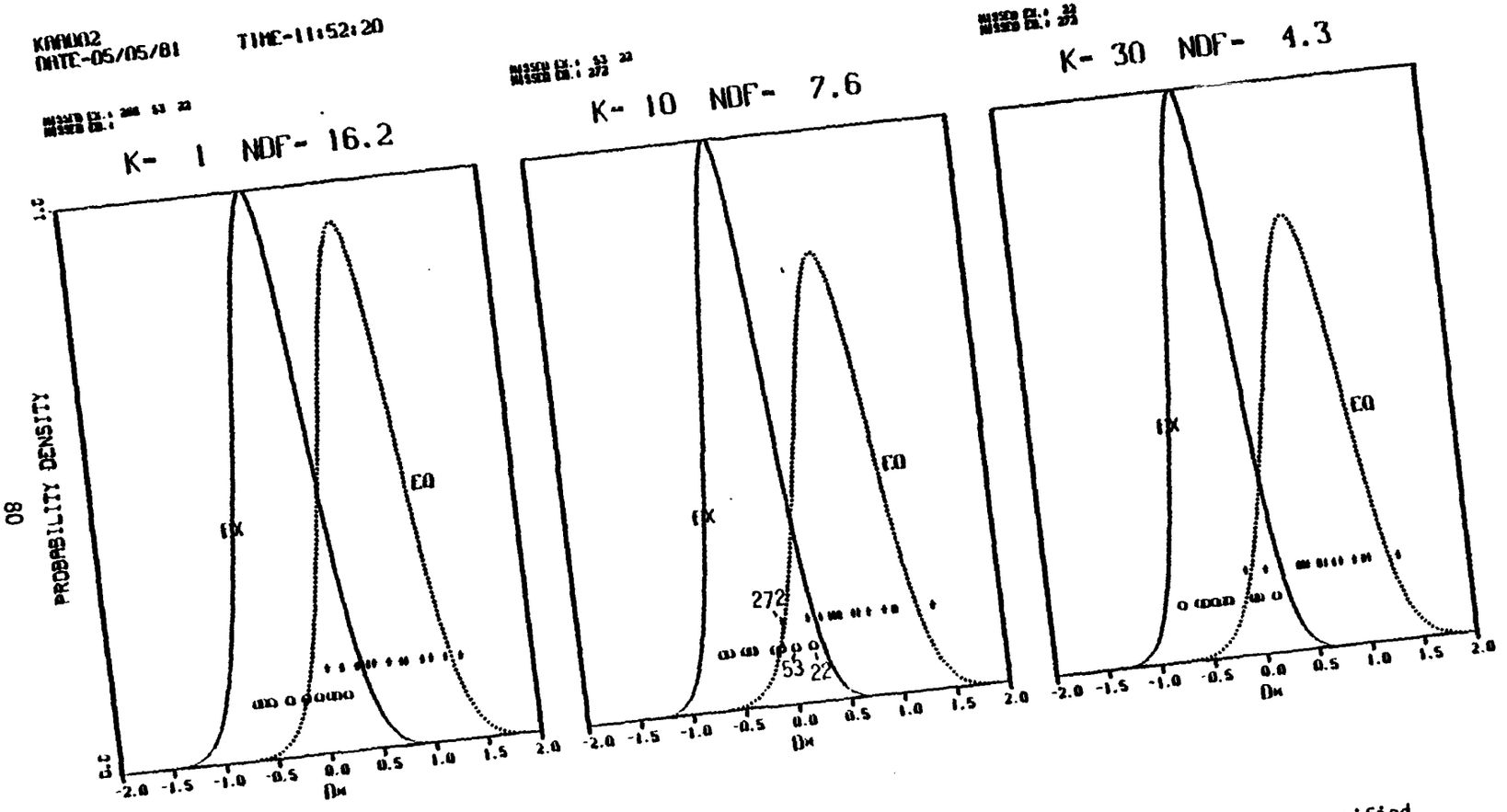


Figure 29. The jackknife calculation for KAAO VFM data shows that the fewest number of misclassified events occurs when NDF = 4.3. This corresponds to a low dimension to the decision hyperplane and a smooth set of feature weights.

the earthquake events by crosses. In order to display graphically the clustering of the population values of d* and their spread, we have superimposed on the univariate plots simple Gaussian functions (the solid line denotes the explosion function, and the dash line the earthquake function) calculated from the usual formula by inserting the mean and standard deviation of the two sets of d* values. There is, of course, no reason to presuppose that the jackknifed values of d* do represent samples from the Gaussian population. If they did, however, and if the earthquake and explosion data sets had equal variances, then the two Gaussian plots would appear to be centered at -0.5 and +0.5 for the earthquakes and explosions respectively, and have equal amplitudes. In this case, for example, we note that for all three values of the damping parameter, the earthquake Gaussian is of somewhat lower amplitude than the explosion Gaussian which is a reflection of the greater spread in the values of d* obtained by jackknifing the earthquake data.

Not only does the jackknife calculation for each value of the damping parameter yield 29 slightly different weight vectors, it also yields 29 different values for the decision constant, b. The heuristic basis for our definition of this constant was mentioned earlier. It is likely that a different definition of b would lead to somewhat different results for the values of d* shown in the tnree panels of this figure. Just as the choice of b amounts to expressing a rule for classification of the linear discriminant, so, in Figure 29, one may select a critical value of d*, say d_c^*, to perform classification. Adopting the rule that an event is classified as an explosion if d* is negative, and classified as an earthquake if d* is positive, i.e., $d_c^* = 0$, then the misclassified events are obtained as shown in the top part of each panel. Note that the set of misclassified events depends upon the choice of the damping parameter. When the damping parameter is large, only a few principal components are retained in the covariance matrix. This leads to a rather smooth set of weights and results in only two misclassifed events for this data set. For smaller values of the

damping parameter, there are a larger number of degrees of freedom, and more events are misclassified. Note, for example, that event 53 is correctly classifed for K = 30 and incorrectly classified for the other two cases, whereas event 272 is correctly classified for 16 degrees of freedom, but misclassified for 7.6 and 4.3.

The direct correlation between degrees of freedom and misclassified events obtained in the jackknife test, is just contrary to the result one obtains when all the data are lumped together. Since higher degrees of freedom amounts to a higher dimension in the discrimination hyperplane, the number of misclassified events decreases as the degrees of freedom increases.

Although we do not wish to attach too much importance to the fit of Gaussian dispersion functions about the two sets of d* values, we note that the value of d* at which the solid curve and dotted curve cross depends upon the value of the damping parameter. For the damping parameter of 30, the intersection point occurs nearly at d* = 0; whereas for K = 1 it is approximately d* = 0.2. This illustrates the phenomenon discussed in the theoretical discussion that there is there is no a priori reason why the Fisher linear discriminant yields equal misclassification probablilities for the two populations.

The principal purpose behind the jackknife study is to provide more realistic estimates of misclassification probabilities than are obtained when the data are treated in toto, and to provide a means for quickly identifying anomalous seismograms. We find, for example, that in this study, events 22, 53, 266, and 272 give ambiguous results, for both d and d* are close to zero. Small changes in the definition of the constant, b, or alternatively of the decision value d_c^* could flop these events into one group or another. Different choices of the damping parameter have the same effect.

Linear discriminate analysis with jackknifing completely supports the conclusion of Savino, et al. (1980a) that the VFM method (spectral ratios) is an effective way of separating the explosion seismograms from the earthquake seismograms in the Kabul

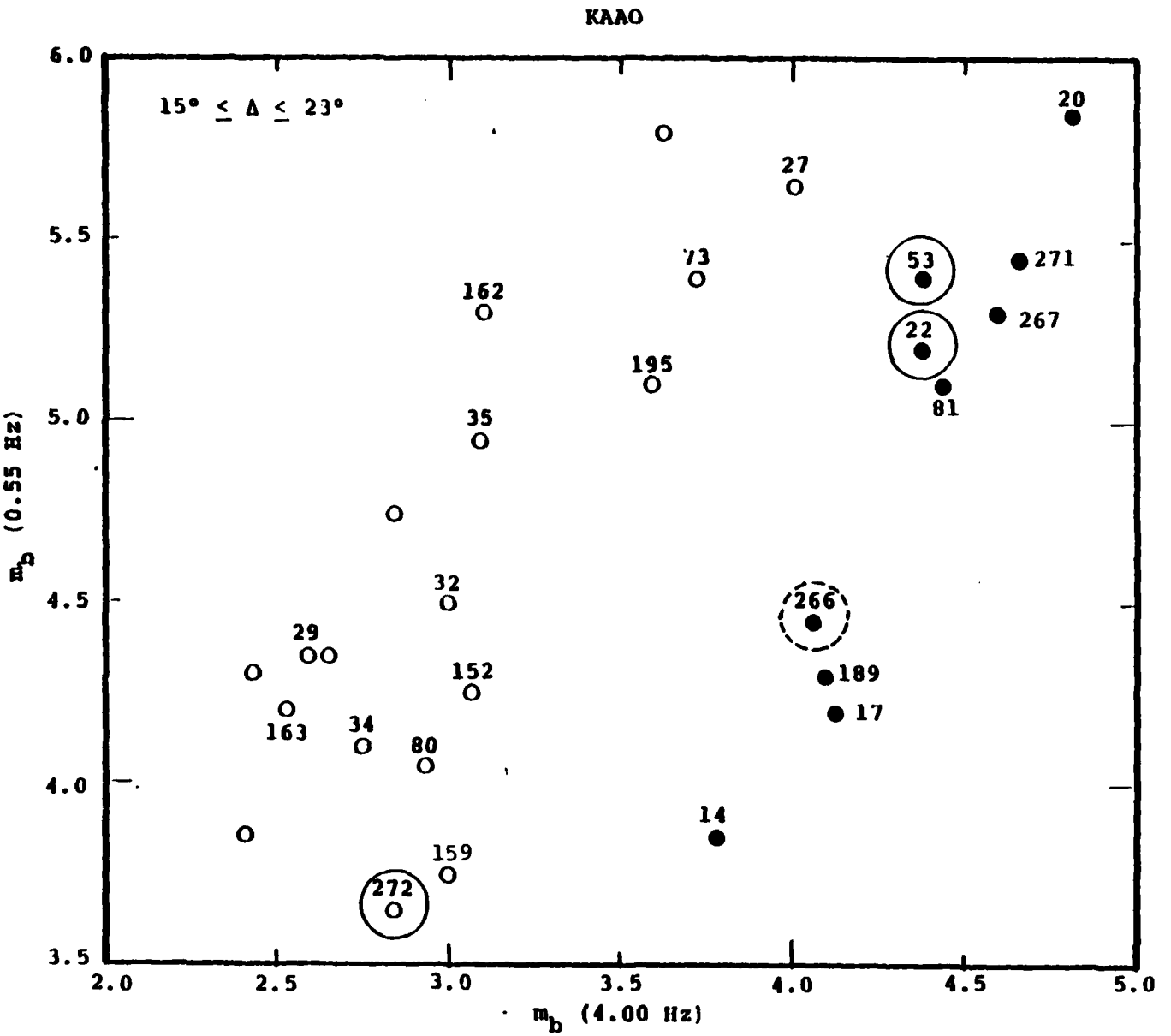


Figure 30. A $m_b(4.0)/m_b(0.55)$ bivariate plot of the VFM data at Kabul (after Savino) shows that the four events (circled) which were misclassified by one or more of the jackknife calculations (see Figure 35) are all on the inner borders of the two event clusters.

data set. This is illustrated in Figure 30 (reproduced from Savino, et al.) which shows the results of the bivariate discriminant analyses of the Kabul data. It can be seen that earthquake 272, and explosions 22 and 53, all indicated by circles, fall on the inner boundaries of their respective bivariate populations, and that explosion 266, which jackknifing missed only for the large degree of freedom case, is also more earthquake like than the seven other correctly classified explosions. Note also on this figure that earthquake 159 lies very near the misclassified earthquake event 272. Table 1 shows that this earthquake, although correctly classified, would have a rather large uncertainty attached to its classification.

The difference between the methodology of Savino and that used here should be mentioned again. The basic data for the two calculations was identical, and that consisted of the spectral magnitude at 40 different frequencies spanning the range 0.5 Hz to 5.0 Hz for 29 events. Our analysis has taken all 40 spectral estimates for tne set of events and found the linear weighting of the relative spectral magnitudes which forms the best separation into two groups. There was an arbitrary parameter in this calculation, the damping parameter, which significantly altered the details of the weight vector, but only slightly altered the final event classifications. This was most clearly shown in the jackknife experiment. Savino, on the other hand, selected just two frequency bands within the range 0.5 Hz to 5.0 Hz. Across each of the frequency bands, a high order polynomial was fitted to the various spectral amplitudes. This interpolating polynomial was then evaluated at a specific frequency to yield the VFM magnitude presented in the bivariate plots.

It does not take much artistic skill to be able to draw a straight line on Figure 30 which totally separates the earthquake population from the explosion population. How is it, then, that the linear discrimination analysis makes mistakes, thus appearing to perform less satisfactorily. The answer is, in fact, that it does

not. This may be seen by the data presented in by Table 1. Column two in this table gives event numbers. For each event, column seven, labeled d(all), gives the value of the scalar discriminant obtained when all the data is processed at once using a damping parameter of 10. Next to the d(all) column is the column labeled "original errors" which tells those events which are misclassified under this criterion. It can be seen that this column is empty. On the other hand, when the data are jackknifed, a slightly different set of scalar discriminant values, our d*, is obtained as shown in column dstar. Now we discover that there have been three misclassified events on the jackknife calculation.

For classifying future events, one wants to take the largest possible training set, and it is the set of weights shown on the K = 10 panel of Figure 28 which are included in the automatic discrimination program. When these weights are applied to the 29 events in the test set, we obtain the values of d(all) shown in Table 1. These values of d(all) roughly correspond to the perpendicular distance between the event data vector and the separation plane. When reduced to two dimensions, d(all) correlates with distances measured on the bivariate plot shown in Figure 30. For example, the most negative value of d(all) was obtained for event 14, and the most positive value was obtained for earthquake 162. If we look at the position of these two events on Figure 30, we see that event 162 is well on the outer boundary of the earthquake population. (Event 14, on the other hand, is more toward the central zone of the explosion population.)

4.3 RKON RESULTS

The RKON data set consists of the 54 events listed in Table 2. The relative m_b plots shown at the bottom of Figure 31 indicate again that the explosions are relatively richer in high frequencies than are the earthquakes. The three weight vectors obtained for the three choices of the damping parameter show the same tendency to become erratic as the damping is decreased and the

TABLE 2

RKON SORTED JACKKNIFE SUMMARY

INDEX	EVENT	DELTA	MBREF	TYPE	CLASS	D-ALL	ORIGINAL ERRORS	DSTAR	JACKKNIFE ERRORS
1	22	76.4	4.84	-1	1	-.71		-.74	
2	19	54.2	5.35	-1	1	-.71		-.70	
3	34	96.8	4.93	1	2	-.42	2	-.70	2
4	20	79.3	5.35	-1	1	-.70		-.68	
5	81	79.4	5.30	-1	1	-.65		-.58	
6	21	71.0	4.88	-1	1	-.60		-.58	
7	77	85.9	4.38	1	2	-.27	2	-.48	2
8	28	90.7	4.56	2	2	-.07	2	-.38	2
9	33	53.8	4.63	-1	1	-.49		-.21	
10	143	59.6	4.96	1	2	.08		-.04	2
11	37	60.3	4.01	1	2	.23		.14	
12	62	66.8	4.09	2	2	.25		.20	
13	60	67.2	4.00	1	2	.24		.22	
14	47	98.7	4.85	1	2	.27		.22	
15	24	96.5	5.20	1	2	.34		.28	
16	4	91.6	4.30	2	2	.39		.35	
17	144	71.7	4.26	2	2	.39		.35	
18	145	79.7	4.23	3	2	.38		.35	
19	50	55.4	5.54	1	2	.41		.36	
20	38	92.0	4.34	3	2	.38		.36	
21	9	91.6	4.69	2	2	.40		.38	
22	35	96.9	4.54	1	2	.40		.38	

TABLE 2 (continued)

INDEX	EVENT	DELTA	MBREF	TYPE	CLASS	D-ALL	ORIGINAL ERRORS	DSTAR	JACKKNIFE ERRORS
23	27	102.8	4.62	1	2	.44		.40	
24	49	64.0	4.48	2	2	.43		.40	
25	68	97.1	4.78	1	2	.42		.41	
26	30	96.8	4.87	1	2	.45		.41	
27	48	92.1	4.21	3	2	.46		.41	
28	29	101.4	4.44	1	2	.43		.42	
29	57	63.3	4.14	1	2	.46		.45	
30	39	96.8	4.45	1	2	.48		.46	
31	32	96.8	4.50	1	2	.47		.46	
32	23	91.5	4.79	1	2	.49		.47	
33	41	91.3	5.11	2	2	.51		.50	
34	36	65.9	4.30	1	2	.51		.50	
35	61	87.1	3.97	1	2	.52		.50	
36	45	92.1	4.22	2	2	.52		.52	
37	46	98.7	4.63	1	2	.54		.52	
38	66	89.0	4.32	1	2	.52		.52	
39	10	92.2	4.17	2	2	.53		.53	
40	3	67.9	4.09	1	2	.55		.55	
41	7	99.8	5.08	1	2	.57		.57	
42	59	89.3	4.06	1	2	.58		.58	

AD-A113 369 SYSTEMS SCIENCE AND SOFTWARE LA JOLLA CA F/G 17/10
AUTOMATIC SEISMIC SIGNAL PROCESSING RESEARCH.(U)
SEP 81 W E FARRELL, J R MURPHY, W L RODI F08606-80-C-0020
UNCLASSIFIED SSS-R-81-5186 VSC-TR-82-17 NL

END
DATE FILMED
5-82
DTIC

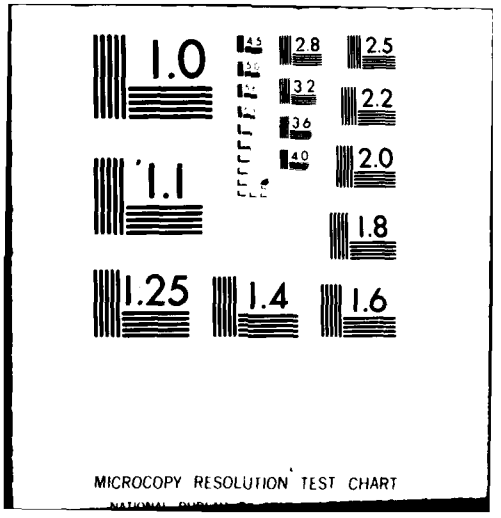

1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
MICROCOPY RESOLUTION TEST CHART

TABLE 2 (continued)

INDEX	EVENT	DELTA	MBREF	TYPE	CLASS	D-ALL	ORIGINAL ERRORS	DSTAR	JACKKNIFE ERRORS
43	65	71.2	4.08	1	2	.60		.60	
44	146	55.9	3.93	1	2	.62		.60	
45	8	80.0	3.81	1	2	.65		.65	
46	76	91.5	4.20	1	2	.67		.67	
47	25	91.4	4.33	1	2	.67		.68	
48	78	101.4	4.34	1	2	.69		.70	
49	55	64.2	4.25	1	2	.73		.74	
50	58	59.1	3.92	1	2	.75		.75	
51	79	79.3	3.95	-1	1	.11	1	.80	1
52	26	92.8	4.25	2	2	.88		.90	
53	56	67.2	4.15	1	2	.90		.93	
54	31	96.9	4.39	1	2	.91		.96	

The summary listing of the jackknife results shows that one event (143) which is properly classified when all the data are analyzed, is misclassified when it is jackknifed.

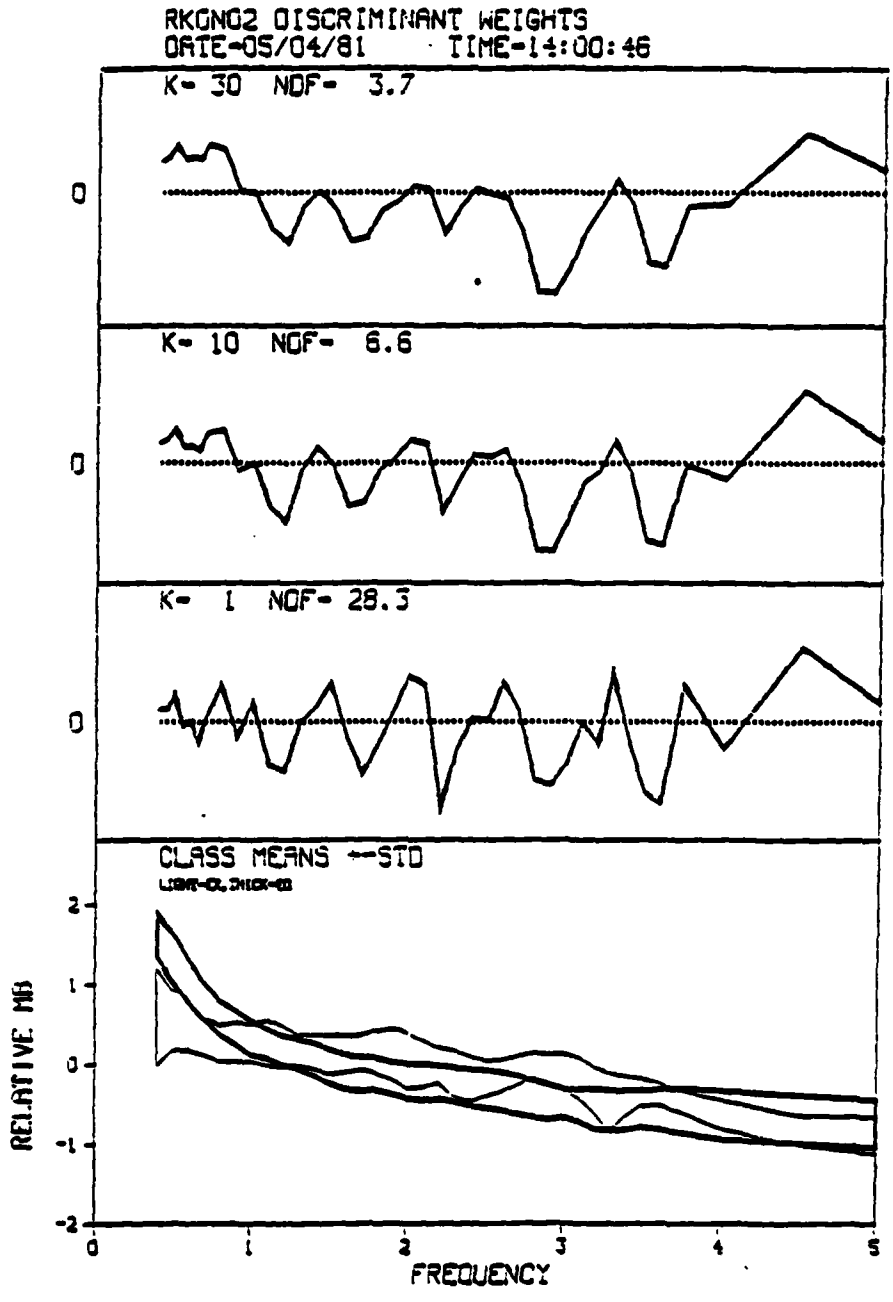


Figure 31. The plot of the feature weights vectors for the VFM discriminant at RKON (three top panels) shows most clearly the separation in the relative magnitudes for the two classes for a small number of degrees of freedom.

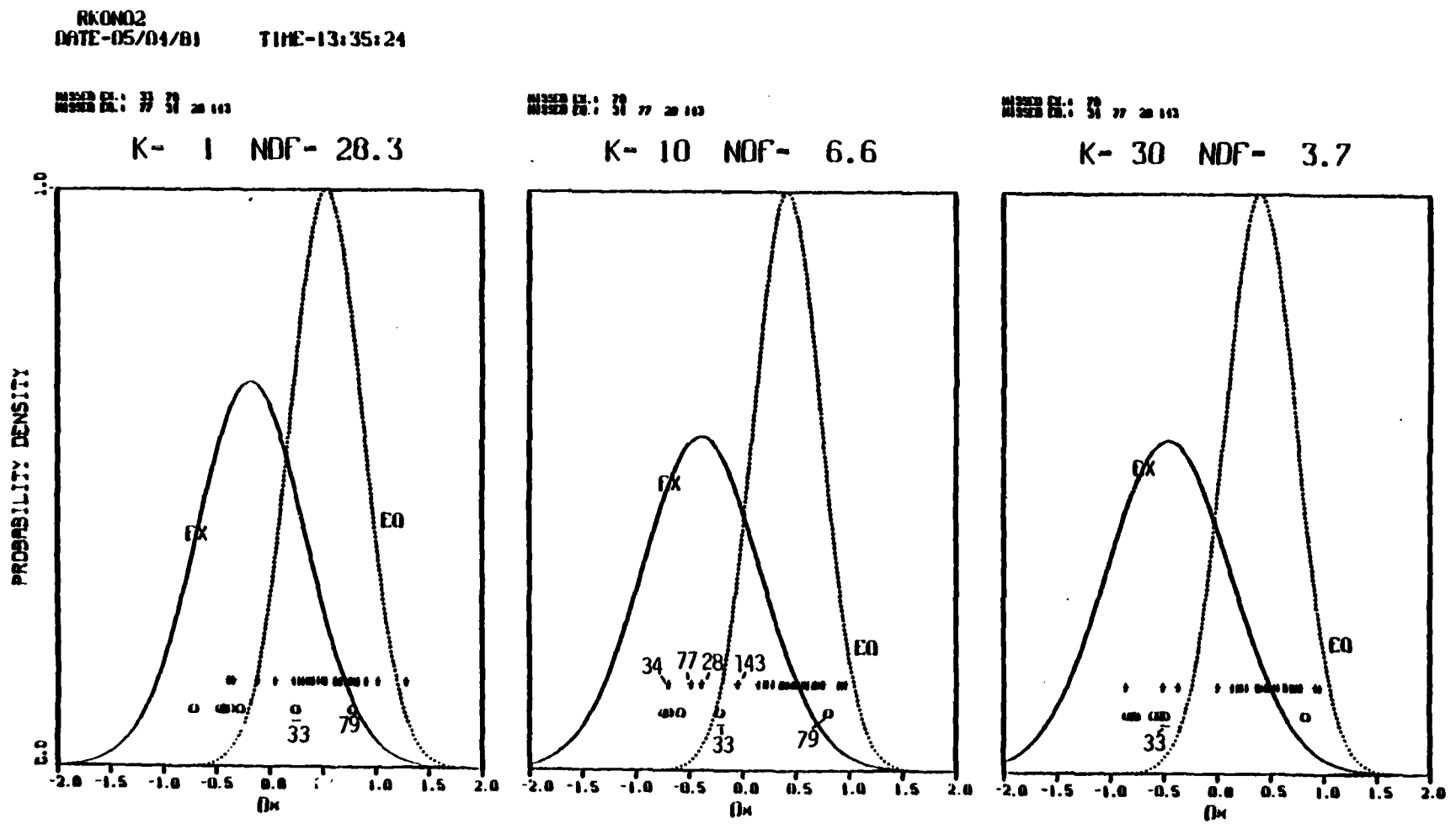


Figure 32. The jackknife calculations for RKON VFM data show that the number of misclassified events does not depend strongly on the number of degrees of freedom.

degrees of freedom increases, and, for the K = 1 case, the low frequency to high frequency slope is almost completely lost. For K = 10 and K = 30, we note that, although for frequencies above 1.0 Hz the weights tend to be negative almost everywhere, there is a pronounced scalloping in the set of weights. The bivariate discrimination used by Savino selected the two frequencies 0.6 Hz and 3.25 Hz. It is amusing to note that there is a pronounced dip in the weight vector around 2.9 Hz, and one speculates that a slight decrease in the upper frequency could possibly have produced an even more disjoint separation of the events in the data set. Again, the set of weights which have been incorporated into the automatic discrimination routine are those obtained for the damping parameter of K = 10.

The jackknife calculation of the RKON data for the three values of the damping parameter (Figure 32) shows that the explosions have a wider spread than the earthquakes. Whereas Kabul produced a somewhat tighter clustering of the values of d* for the explosion population, here we find exactly the opposite -- that the tightest grouping of events occurs for the d*'s obtained for the earthquake population. However, if one looks in detail at the values of the discriminate function d*, we see that the wide spread in tne explosion population is determined principally by two events, event 33 and event 79. Event 79 is so clearly anomalous that it would certainly be better to recalculate this example leaving event 79 out of the data set. That would have the effect of producing an explosion Gaussian which was much less broad.

The tendency of d* for some events to be affected by the choice of damping parameter (previously noted in the discussion of tne Kabul data) is particularly pronounced for event 33 at RKON. We see that when the damping parameter is small (the number of degrees of freedom is large), event 33 is clearly misclassified. However, as the damping is increased, the set of weight becomes smoother, and event 33 moves appreciably to the left. For the largest damping (that is, tne fewest degrees of freedom), event 33 is indistinguishable from its neighbors in the explosion population. The right most

panel in Figure 32 shows particularly clearly the fact that the Gaussian width of the explosion population is entirely controlled by the single anomalous event, explosion number 79.

As before, there is a clear relationship between the results obtained in this analysis based upon the linear discriminant function and the jackknife calculation and the bivariate discrimination of Savino, et al. (1981a). This is shown in Figure 33. Three events, explosion 79 and earthquakes 28 and 34, are misclassified by both criteria. Explosion 33, which was missed only when the damping parameter was set anomalously low, and earthquakes 77 and 143, are positioned in the ambiguous central portion of the bivariate plot. We note again, however, that there are some events, in this case earthquake 7 and earthquake 24, which are correctly classified by the linear discriminant analysis whereas a neighbor, earthquake 143, is misclassified for all three values of a damping parameter.

If one were to draw a line on Figure 33 which best separated the two populations, the three events, 79, 28 and 34 would be incorrectly classified. Table 2 (see column Original Errors) indicates that the linear discriminant analysis of all the data misclassified these three events also, as well as one additional event, earthquake number 77. Jackknifing changes the picture only slightly by indicating the marginal nature of event 143.

4.4 ILPA RESULTS

The VFM results for the Iranian long period array have been selected for the final presentation of linear discriminant analysis and jackknifing. Although there are 56 events in the data set (see Table 3), only four of these are explosions. Figure 34, at the bottom, shows again the band of magnitudes which encompasses 95 percent of the relative m_b's for the explosion and earthquake classes. The three panels at the top of this figure show the three sets of weight vectors obtained from the linear discriminant

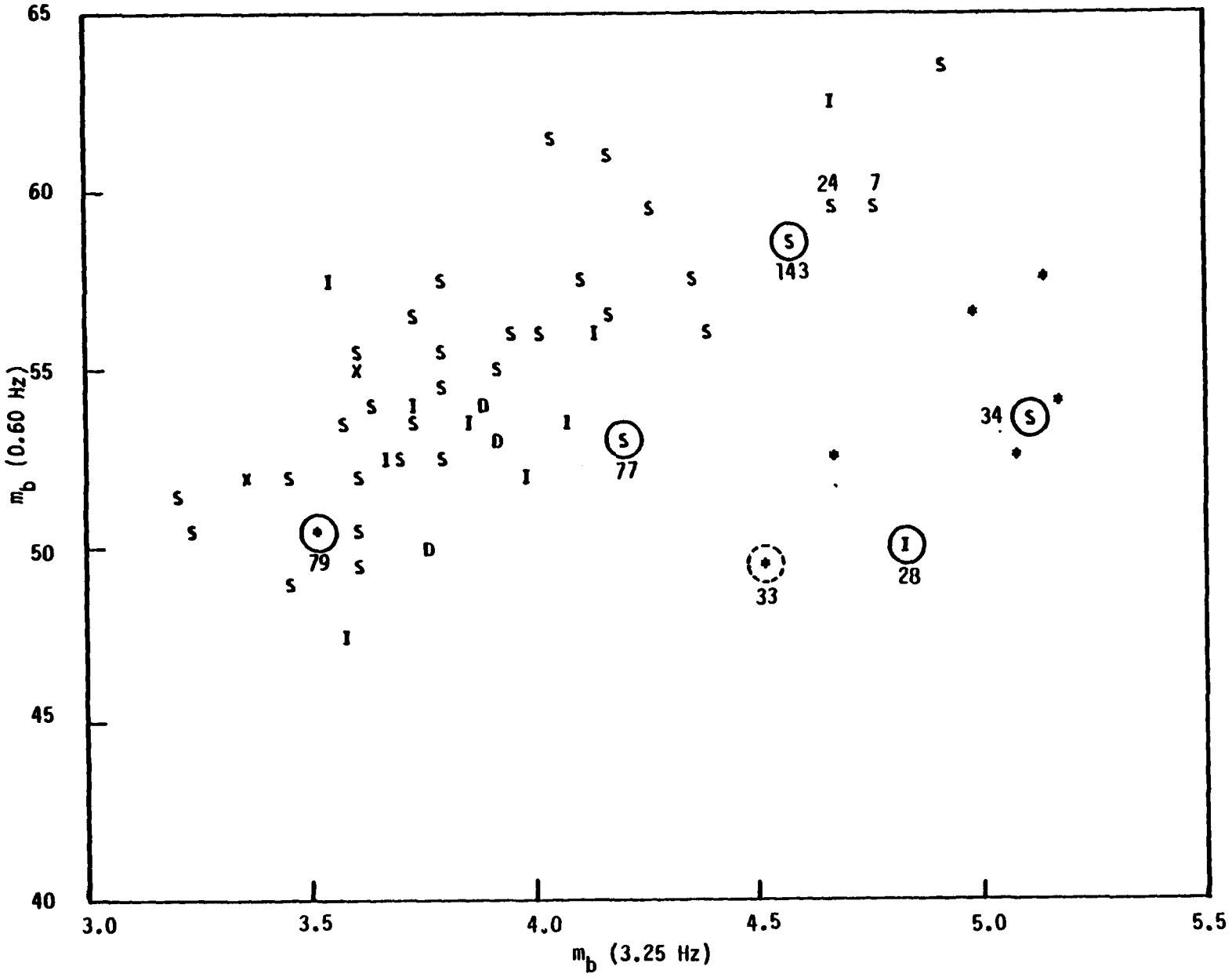


Figure 33. The $m_b(3.25)/m_b(0.60)$ bivariate plot of the VFM data at RKON (after Savino) shows that three out of the five events misclassified by the jackknife procedure fall well within the wrong population cluster. Two shallow earthquakes (77 and 143) are borderline cases.

TABLE 3

ILPA SORTED JACKKNIFE SUMMARY

INDEX	EVENT	DELTA	MBREF	TYPE	CLASS	D-ALL	ORIGINAL ERRORS	DSTAR	JACKKNIFE ERRORS
1	147	68.1	5.10	3	2	-.55	2	-.73	2
2	33	38.3	4.00	-1	1	-.58		-.53	
3	18	41.1	4.30	-1	1	-.56		-.42	
4	19	38.1	5.19	-1	1	-.49		-.35	
5	166	38.5	4.21	1	2	.00		-.33	2
6	65	43.9	4.22	1	2	.12		-.14	2
7	145	61.4	4.35	3	2	.17		.09	
8	27	37.2	4.09	1	2	.20		.13	
9	29	34.1	4.02	1	2	.29		.26	
10	39	31.3	4.32	1	2	.31		.27	
11	63	69.1	4.44	3	2	.33		.29	
12	31	31.3	4.08	1	2	.34		.31	
13	169	71.7	4.42	2	2	.36		.33	
14	143	72.7	4.82	1	2	.40		.34	
15	34	31.2	4.12	1	2	.39		.34	
16	178	72.7	4.57	1	2	.40		.35	
17	57	70.2	4.40	1	2	.38		.36	
18	193	72.4	4.02	1	2	.39		.36	
19	6	45.2	4.07	1	2	.39		.36	
20	32	31.3	4.02	1	2	.44		.43	
21	3	56.2	4.16	1	2	.45		.43	

TABLE 3 (continued)

INDEX	EVENT	DELTA	MBREF	TYPE	CLASS	D-ALL	ORIGINAL ERRORS	DSTAR	JACKKNIFE ERRORS
22	25	32.2	4.23	1	2	.47		.45	
23	180	72.6	4.16	1	2	.47		.45	
24	183	74.3	4.28	1	2	.52		.49	
25	144	69.8	4.25	2	2	.51		.50	
26	64	72.3	4.35	2	2	.51		.50	
27	165	73.3	4.25	1	2	.52		.51	
28	47	72.5	3.76	1	2	.53		.51	
29	46	35.9	3.92	1	2	.53		.52	
30	146	72.3	4.19	1	2	.53		.53	
31	24	31.6	4.16	1	2	.54		.53	
32	21	42.7	3.93	-1	1	-.08	1	.53	1
33	186	72.2	4.27	1	2	.55		.54	
34	171	72.8	4.17	1	2	.56		.55	
35	75	31.2	4.16	1	2	.62		.61	
36	195	32.2	3.84	1	2	.63		.61	
37	157	72.0	4.23	2	2	.64		.63	
38	192	34.1	3.71	1	2	.65		.63	
39	164	70.7	4.82	1	2	.68		.65	
40	176	70.6	4.16	1	2	.68		.67	
41	68	30.8	3.94	1	2	.69		.69	
42	159	30.6	4.08	1	2	.71		.71	

TABLE 3 (continued)

INDEX	EVENT	DELTA	MBREF	TYPE	CLASS	D-ALL	ORIGINAL ERRORS	DSTAR	JACKKNIFE ERRORS
43	156	72.2	4.33	1	2	.76		.76	
44	58	72.5	4.35	1	2	.80		.81	
45	78	34.1	4.01	1	2	.81		.81	
46	56	73.1	4.36	1	2	.81		.82	
47	37	61.3	4.20	1	2	.83		.83	
48	182	72.7	4.26	1	2	.86		.86	
49	77	52.4	4.45	1	2	.87		.90	
50	194	71.9	3.71	1	2	.89		.90	
51	50	72.6	4.81	1	2	.90		.92	
52	55	72.9	4.22	1	2	1.00		1.01	
53	187	32.6	4.20	1	2	1.02		1.04	
54	30	31.2	4.68	1	2	1.00		1.05	
55	160	62.2	4.52	1	2	1.09		1.11	
56	154	68.9	4.58	3	2	1.27		1.33	

The summary listing of the jackknife results shows that two earthquakes and one explosion which are properly classified when all the data are analyzed are misclassified when it is jackknifed.

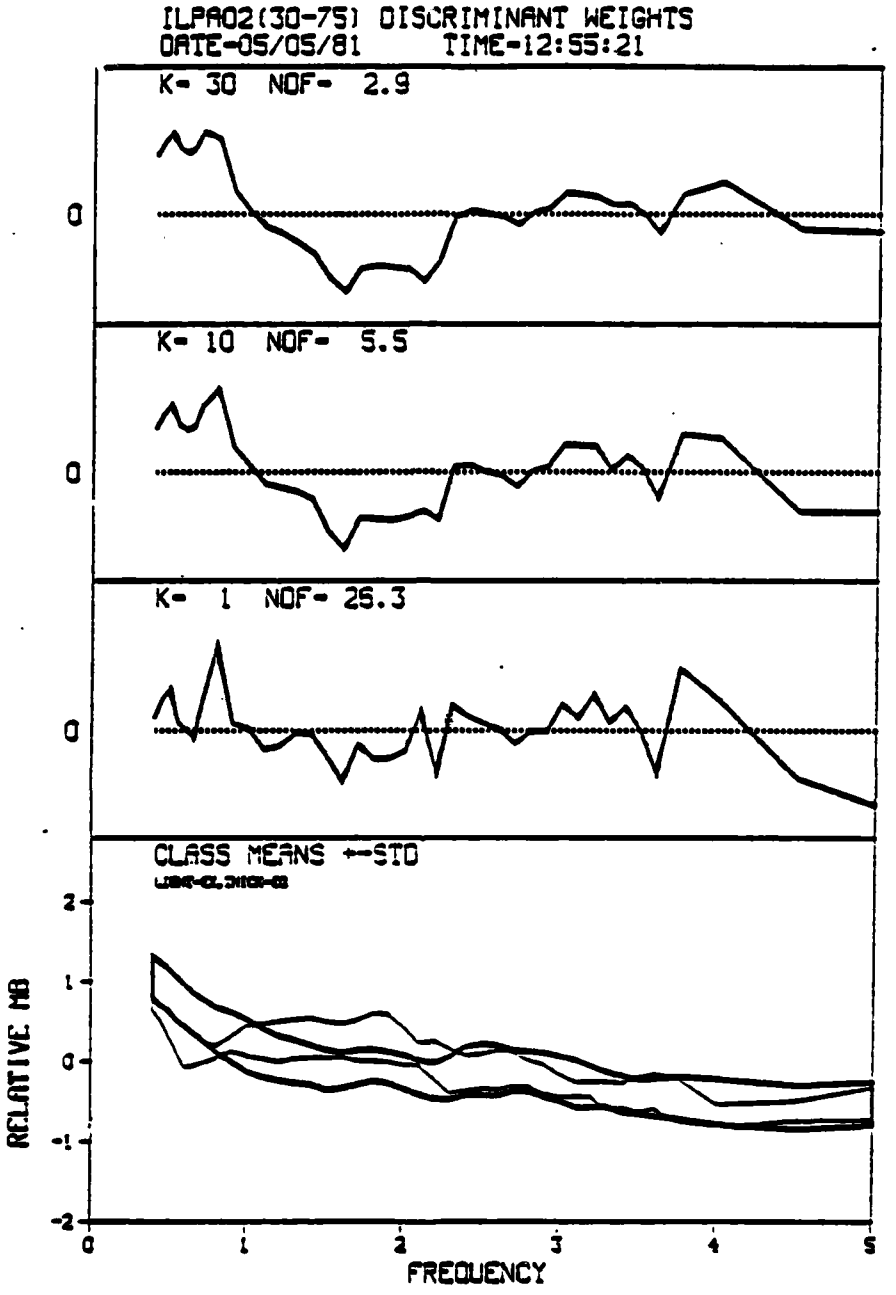


Figure 34. The plot of the feature weights vector for the VFM discriminant at ILPA clearly reflects the separation between the two population mean magnitudes over the range from 0.5 to 2.0 Hz.

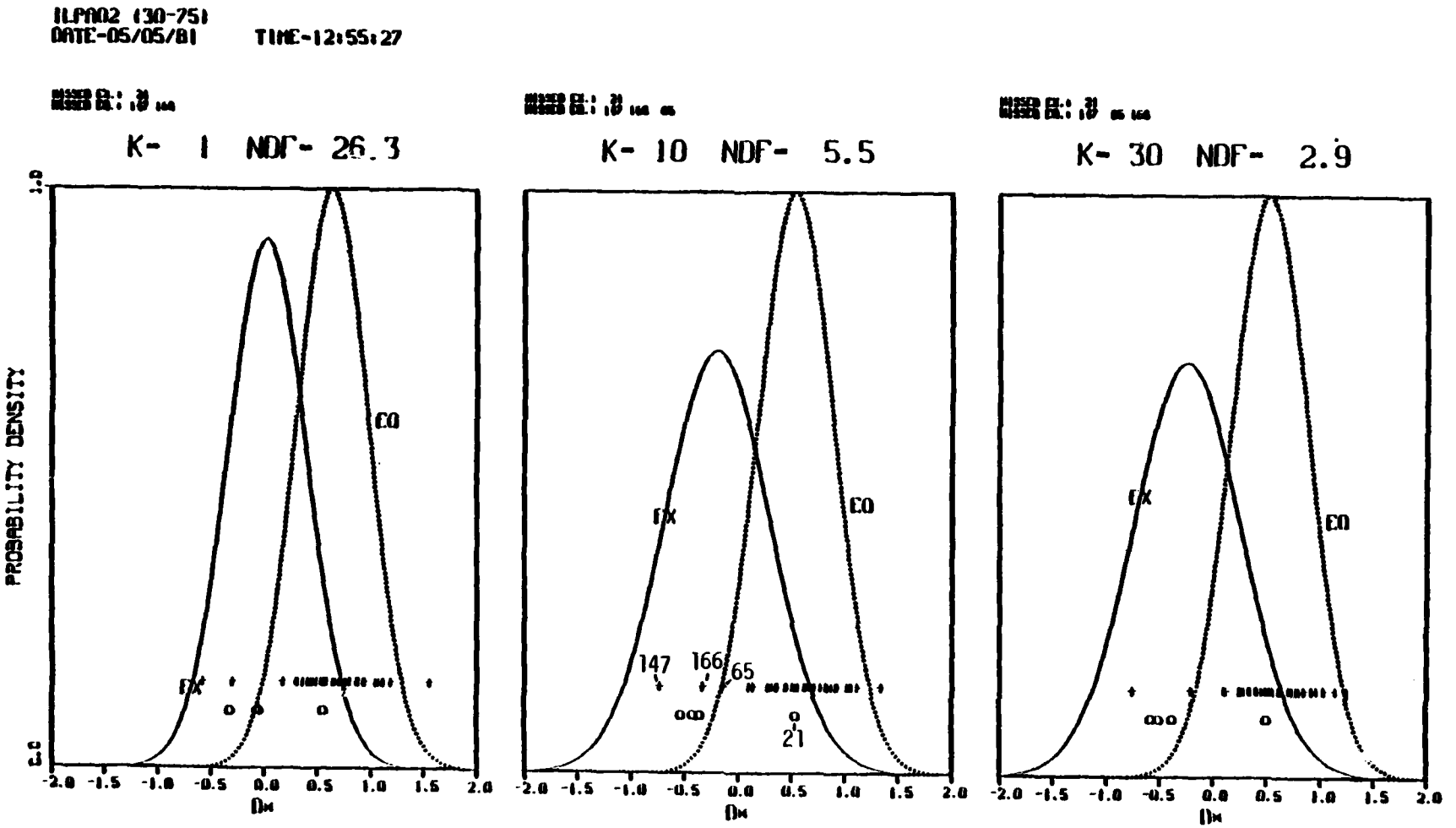


Figure 35. Jackknifing shows that three misclassified events are anomalous for all three choices of the number of degrees of freedom.

analysis. We note here that, since the data set is larger, the numbers of degrees of freedom for the three choices of damping parameter is somewhat larger than it was in the past. The tendency of the weights to become more jagged as the damping decreases is again apparent as is the positive to negative trend with increasing frequency. Note that for all three values of the damping parameter, the weight vector components are essentially zero for frequencies above 2.5 Hz, and this clearly mimics the lack of separation between the two relative m_b classes shown in the bottom panel. Savino selected for the bivariate discrimination, the frequencies 0.55 Hz and 2.0 Hz. If one were to select a single pair of frequencies for performing discrimination from the plots of the weight vectors shown here, one might prefer the choice 0.75 and 1.75 Hz. Whether or not this would yield an improvement in Savino's method is not known.

When the linear discriminant function is jackknifed for the three choices of the damping parameter, the results presented in Figure 35 are obtained. Explosion 21 is missed in all three cases, and its d* value of 0.5 clearly places it in the midst of the earthquake population. The paucity of explosion data causes the best fitting Gaussian to change radically as the damping parameter is changed. As with the RKON measurments, a single anomalous event, event 21, has caused the explosion Gaussian to be significantly broader than the earthquake Gaussian. With only four explosion datums, and with one of them clearly being anomalous, it is obvious that event 21 is having an undue influence on the choice of linear discriminant weights, and these data should probably be reanalyzed, deleting event 21 from the analysis. Just as explosion 21 is misclassified for all three choices of the damping parameter, so earthquakes 147 and 168 are misclassified for all three choices. Event 65 is a borderline case which is correctly classified when the damping is strong, but misclassified when the damping is small. We recall that when the damping is small, the number of degrees of freedom is high, and one is fitting a higher dimensioned hyperplane through the parameter space. It is to be expected that when the damping is small, one will have a smaller number of misclassified

events, but that is inevitably accompanied by an increased uncertainty in the classifications of every event.

The comparison between the linear discriminant and jackknife classification methods and the bivariate classification performed by Savino, et al. is presented in Figure 36. Explosion 21 and earthquake 147 clearly fall into the wrong groupings from both points of view. Events 65 and 166, which are misclassified in a linear discriminant analysis, were not included in the bivariate discrimination by Savino. The comparison between the value of the linear discriminant d(all) obtained when the entire data set is processed with a damping parameter of 10 and the set of values d* obtained for the linear discriminant when the data is jackknifed is shown in Table 3. As was noted in Figure 36, a linear discriminant classification obtained by weighting all frequencies is no more effective than a bivariate discriminant based on the frequencies of 0.55 Hz and 2.0 Hz in classifying the earthquake 147 and the explosion 21. It can be seen in the right most column that event 65 and 166, which were not discussed by Savino, are clear borderline cases which are correctly classified when the data is lumped together, but which are misclassified when the jackknife is performed.

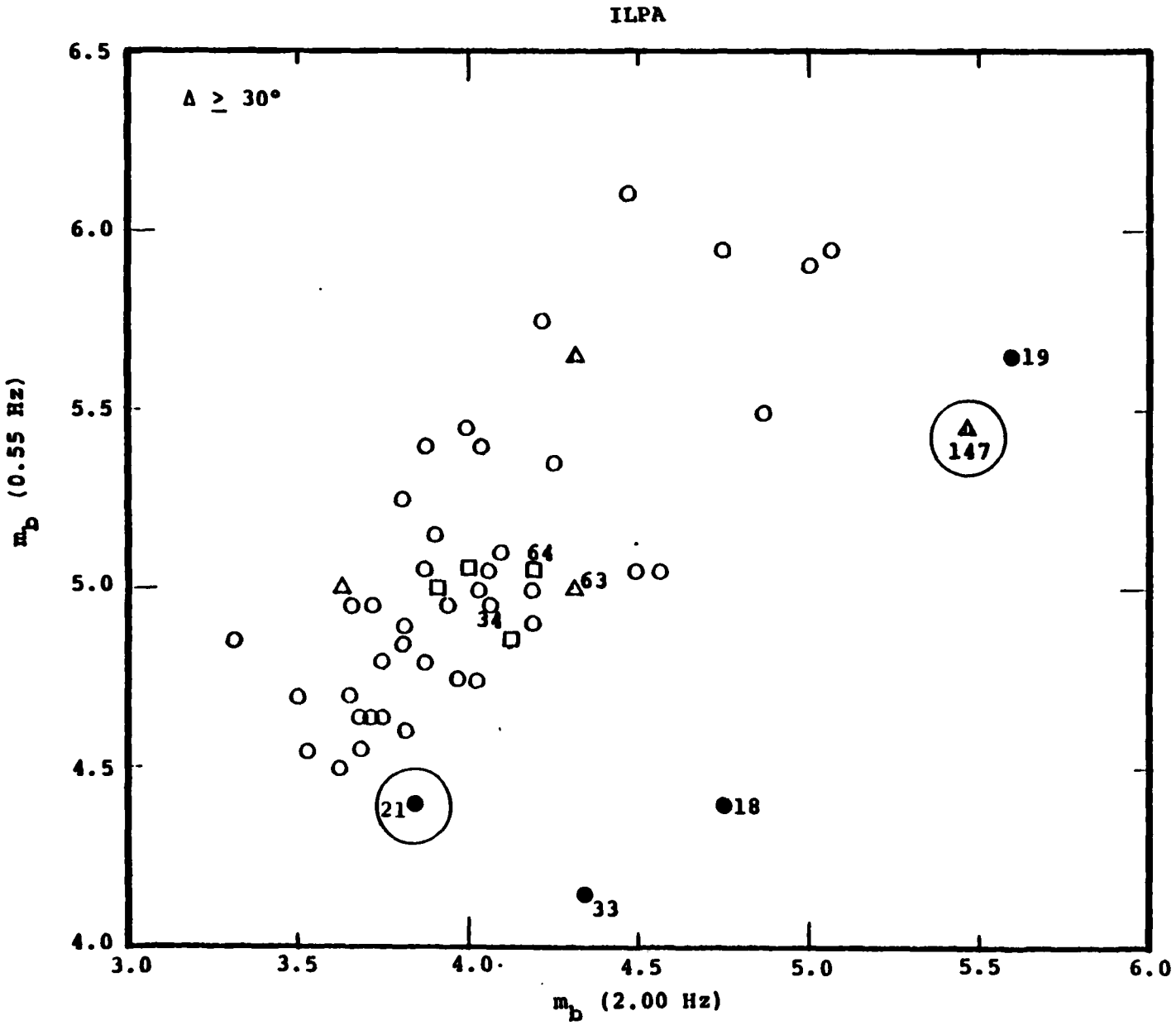


Figure 36. The $m_b(2.0)/m_b(0.55)$ bivariate plot of the VFM data at ILPA shows that three of the misclassified events are well removed from their respective population means.

5. ACKNOWLEDGMENTS

R. Blandford, W. Rivers, and D. Von Seggern of Teledyne-Geotech have been most courteous in discussing seismic discrimination with us. R. Romine gave much assistance as we learned about Unix. Several extended conversations with W. Nicholson of Battelle have brought many statistical matters into sharper focus and have turned us away from several blind alleys. B. Barker of VSC played an important role in supervising the work and providing us with data.

6. REFERENCES

Anderson, t. W. (1958), An Introduction to Multivariate Statistical Analysis, Wiley, New York.

Bache, T. C., W. J. Best, R. R. Blandford, G. U. Bulin, D. G. Harkrider, E. J. Herrin, A. Ryall, and M. J. Shore (1981), "A Technical Assessment of Seismic Yield Estimation," Defense Advanced Research Projects Agency Report DARPA NMR 81-02.

Bache, T. C., J. R. Murphy, S. M. Day, T. J. Bennett, and B. Shkoller (1980), "Regional Detection of Decoupled Explosions, Yield Estimation From Surface Waves, Two-Dimensional Source Effects, Three-Dimensional Earthquake Modeling and Automated Magnitude Measures," Systems, Science and Software Semiannual Technical Report SSS-R-80-4594, Submitted to AFTAC/VSC, July.

Chiburis, E. F., R. O. Ahner, E. C. Rienhardt and S. J. Price (1980), "A Preliminary Study of Measuring M and m_b in the Time and Frequency Domain," Ensco, Inc. Report DCS-STR-80-43.

Dahlman, O., and H. Israelson (1977), Monitoring Underground Nuclear Explosions, Elsevier, Amsterdam, Netherlands.

Farrell, W. E. (1981), "Automatic Seismic Discrimination System," Systems, Science and Software Semiannual Technical Report SSS-R-82-5246 (Draft), Submitted to AFTAC/VSC, November.

Farrell, W. E., J. Wang, C. B. Archambeau and R. C. Goff (1980), "Evaluation of the MARS Seismic Event Detector," Systems, Science and Software Report SSS-R-81-4656, August.

Gnanadesikan, R. (1977), Methods for Statistical Analysis of Multivariate Observations, Wiley, New York.

Mosteller, F. and J. W. Tukey (1977), Data Analysis and Regression, Addison-Wesley Publishing, Reading, Massachusetts.

Patrick, E. A. (1972), Fundamentals of Pattern Recognition, Prentice-Hall, Englewood Cliffs, New Jersey.

Rao, C. R. (1973), Linear Statistical Inference and Its Applications, John Wiley and Sons, New York.

Rivers, D. W., M. E. Marshall, J. A. Burnetti, J. A. Wagner, P. J. Klauda and A. O'Donnell (1980), "An Evaluation of the VSC Discrimination Experiment (Draft) (U)," VELA Seismological Center Report TR-81-12 (S) .

Rivers, D. W., D. H. von Seggern, H. S. Sproules, B. L. Elkins and G. R. Naylor (1979a), "A Statistical Discrimination Experiment for Eurasian Events Using a Seventeen Station Network," SDAC Report No. TR-79-1 Teledyne Geotech, Alexandria (S).

Rivers, D. W., D. H. von Seggern, B. L. Elkins and H. S. Sproules (1979b), "A Statistical Discrimination Experiment for Eurasian Events Using a Twenty-Seven-Station Network," SDAC Report No. TR-79-5, Teledyne Geotech, Alexandria, Virginia.

Rivers, D. W., D. H. von Seggern, I. Megyesi, J. A. Burnetti and P. J. Klouda (1979c), "A Statistical Discrimination Experiment for Eurasian Events Using a Thirty-Five-Station Network," SDAC Report No. TR-79-2 Teledyne Geotech, Alexandria, Virginia, (S).

Savino, J. M., C. B. Archambeau and J. F. Masso (1980a), "Discrimination Results from the Priority 2 Stations," Systems, Science and Software Report SSS-CR-80-4566, Submitted to the Advanced Research Projects Agency, (S).

Savino, J. M., C. B. Archambeau and J. F. Masso (1980b), "VFM Discrimination Results for Eurasian Events Using the Priority 1 and Priority 2 Stations, Systems, Science and Software Report SSS-CR-80-4570, Submitted to the Advanced Research Projects Agency (S).

Savino, J. M., J. F. Masso and C. B. Archambeau (1979), "Discrimination Results from the Priority 1 Stations," Systems, Science and Software Report SSS-CR-79-4026, Submitted to the Advanced Research Projects Agency (S).

Sax, R. L., A. G. R. Bell and D. L. Dietz (1979a), "Event Identification Experiment: Priority I Data Set," Report No. SAR(01)-TR-79-01, ENSCO, Inc., Springfield, Virginia.

Sax, R. L., A. G. R. Bell and D. L. Dietz (1979b), "Event Identification Experiment: Combined Priority I/Priority II Data Sets," Report No. SAR(01)-TR-79-07, ENSCO, Inc., Springfield, Virginia.

Tjøstheim, D. (1981), "Multidimensional Discrimination Techniques — Theory and Application," in 'Identification of Seismic Sources - Earthquake or Underground Explosion,' Edited by E. S. Husebye and S. Mykkeltveit, D. Reidel, Holland.

Young, T. Y. and T. W. Calvert (1974), Classification, Estimation, and Pattern Recognition, American Elsevier, New York.

von Seggern, D. H. (1977), "Methods of Automating Routine Analysis Tasks in Preparing a Global Seismic Bulletin," Teledyne Geotech Final Report SDAC-TR-77-13.

Veith, K. F., and G. E. Clawson (1972), "Magnitude From Short Period P-Wave Data," BSSA, 62, pp. 435–452.


DAT
FILM